

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS**

Ernesta Koženiauskaitė

**PANAŠIŲ ONTOLOGIJOS INDIVIDŲ ATPAŽINIMO IR
SUJUNGIMO GALIMYBIŲ TYRIMAS**

Baigiamasis magistro projektas

Vadovė
doc. dr. R. Butkienė

KAUNAS, 2015

**KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS**

**PANAŠIŲ ONTOLOGIJOS INDIVIDŲ ATPAŽINIMO IR
SUJUNGIMO GALIMYBIŲ TYRIMAS**

Baigiamasis magistro projektas
Informacinių sistemų inžinerijos studijų programa (kodas 621E15001)

Vadovė

doc. dr. R. Butkienė
2015-05-25

Recenzentas

doc. dr. S. Drąsutis
2015-05-25

Projektą atliko

Ernesta Koženiauskaitė
2015-05-25



KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS

(Fakultetas)

Ernesta Koženiauskaitė

(Studento vardas, pavardė)

Informacinių sistemų inžinerijos studijų programa, 621E15001

(Studijų programos pavadinimas, kodas)

Baigiamojo projekto „Panašių ontologijos individų atpažinimo ir sujungimo galimybių tyrimas“

AKADEMINIO SAŽINGUMO DEKLARACIJA

20 15 m. gegužės 25 d.
Kaunas

Patvirtinu, kad mano, **Ernestos Koženiauskaitės**, baigiamasis projektas tema „Panašių ontologijos individų atpažinimo ir sujungimo galimybių tyrimas“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Koženiauskaitė, E. "Research on recognition and merging of similar individuals of the ontology". *Final Degree Project of Master of Information Systems Engineering* / Supervisor doc. Rita Butkienė; Kaunas University of Technology, Faculty of Informatics.

Kaunas, 2015. 66 p.

SUMMARY

Every day in websites, news portals the huge amount of information is often redundant. In a multitude of information it is very easy to get lost and not find the necessary information. Automatically analysing websites, news portals, published Lithuanian articles face with a problem that is inserted into the ontology recurring information. In semantically annotated texts, find the necessary information would be much easier and more convenient if the ontology would be given the opportunity to automatically identify similar ontology individuals and link them by similarity.

In this master's degree work an ontology is analysed which was created by Informatics Systems Department during "Syntactic-semantic analysis and search system for Lithuanian Internet, corpus and public sector applications „SemantikaLT“" carried out by Lithuanian Economics growth actions program 3rd priority "Information Society for Everyone" realization means Num. VP2-3.1-IVPK-12-K "Lithuanian Language in Information Society". This ontology individuals of Person class which were automatically created by analysing article texts in Lithuanian language published in the Internet. Creating ontology Persons individuals this way, duplicated Persons were created too.

The master goal is to make possible to automatically identify similar ontology individuals, to link them and thereby minimize information of ontology duplication.

Master thesis objective algorithm was developed and written in SPARQL query that includes the ability to automatically identify similar ontology individuals and link them. This helps to reduce the duplication of ontology information. About specimen collected and analysed information the result is displayed on the ontology that searching every time do not need to analyse the ontology individual's similarity.

Created solution easily adoptable for this type of ontology. In ontology structure there is label_lemma (a generic form of the word), using label_lemma reduced ontology individuals repetitions, merging individuals by similarity for the purpose to provide concrete, structured information.

TURINYS

Lentelių sąrašas	7
Paveikslų sąrašas	8
Terminų ir santrumpų žodynas	10
Įvadas	12
1. Panašių ontologijos individų atpažinimo ir sujungimo galimybių tyrimo probleminės srities analizė	14
1.1. Analizės tikslas	14
1.2. Tyrimo objektas, sritis ir problema	14
1.3. Ontologijos ir jos individų analizė	17
1.4. Tyrimo objekto naudotojų analizė	20
1.5. Panašių ontologijos individų atpažinimo ir sujungimo galimybių analizė	20
1.5.1. Žodžių morfologinės savybės	22
1.6. Darbo tikslas, uždaviniai ir siekiami privalumai	23
1.7. Siekiamo sprendimo apibrėžimas	24
1.8. Analizės išvados.....	25
2. Panašių ontologijos individų atpažinimo ir sujungimo reikalavimų specifikacija ir projektas, formalus aprašas.....	26
2.1. Reikalavimų specifikacija	26
2.2. Dalykinės srities modelis	27
2.3. Reikalavimų apibendrinimas.....	29
3. Panašių ontologijos individų atpažinimo ir sujungimo algoritmas	30
3.1. Sistemos architektūra	32
3.1.1. Reikalavimų analizė.....	33
3.1.2. Loginė visos sistemos architektūra	33
3.1.3. Naudotojo sąsajos klasių modelis	33
3.1.4. Veiklos logikos klasių modelis	34
3.2. Sistemos elgsenos modelis.....	34
3.3. Realizacijos modelis	37
3.3.1. Programinių komponentų architektūra	37
3.3.2. Diegimo modelis	39
4. Sprendimo realizacija ir testavimas	40
4.1. Sprendimo realizacijos ir veikimo aprašas.....	40
4.2. Testavimo modelis, duomenys, rezultatai	42
5. Eksperimentinis sprendimo (Panašių ontologijos individų atpažinimo ir sujungimo) tyrimas	46
5.1. Eksperimento planas	46
5.2. Eksperimento rezultatai.....	46
5.3. Sprendimo taikymo rekomendacijos.....	64
6. Rezultatų apibendrinimas ir išvados	65

7. Literatūra.....66

LENTELIŲ SĄRAŠAS

2.1 lentelė. Panaudojimo atvejo „Sulieti panašius individus, pasirodančius skirtinguose dokumentuose“ specifikacija	26
2.2 lentelė. Nefunkcinis reikalavimas 1	26
2.3 lentelė. Nefunkcinis reikalavimas 2	27
2.4 lentelė. Nefunkcinis reikalavimas 3	27
5.1 lentelė. Žmogaus ir teksto anotatorių rastų asmenų paminėjimų straipsniuose lyginamoji lentelė	60
5.2 lentelė. Žmogaus ir algoritmo rastų ir apjungtų panašių ontologijos individų lyginamoji lentelė	62
5.3 lentelė. Algoritmo teisingai, klaidingai nustatytų ir nesurastų panašumų lentelė.....	63

PAVEIKSLŲ SĄRAŠAS

1.1 pav. Ontologijos užpildymo individualiais veiklos procesas.....	15
1.2 pav. Teksto fragmentai (straipsniai iš internetinio portalo <i>www.delfi.lt</i>) su pasikartojančiais egzemplioriais	16
1.3 pav. Svarbiausios ontologijų rūšys [3].....	17
1.4 pav. Įmonių inžinerijos ontologija [6].....	18
1.5 pav. Individų ontologijos struktūroje aprašymas	19
1.6 pav. Teksto fragmentas (straipsnis iš internetinio portalo <i>www.delfi.lt</i>) su klaidingai parašyta asmens pavarde	21
1.7 pav. Tikslų modelis	24
1.8 pav. Veiklos konceptai	24
1.9 pav. Veiklos kontekstas.....	25
1.10 pav. Veiklos panaudojimo atvejų diagrama	25
2.1 pav. Kompiuterizuojamų PA diagrama.....	26
2.2 pav. Ontologijos meta modelis.....	28
2.3 pav. Dalykinės srities modelis	28
2.4 pav. Asmenų klasę detalizuojantis modelis	29
3.1 pav. Suliejimo pagal vardą ir pavardę, kai pavardė pilna, o vardas sutrumpintas algoritmas	31
3.2 pav. Suliejimo, kai sutampa individų vardai ir pavardės, algoritmas	32
3.3 pav. Analizės (robastiškumo) diagrama.....	33
3.4 pav. Loginė visos sistemos architektūra	33
3.5 pav. Naudotojo sąsajos klasių modelis	34
3.6 pav. Veiklos logikos klasių modelis	34
3.7 pav. Panašių individų suradimas.....	35
3.8 pav. Panašių individų suliejimas.....	36
3.9 pav. Veiklos diagrama, naudojant būsenų mašinos elementus	37
3.10 pav. Programinių komponentų architektūra.....	37
3.11 pav. Komponentų realizacija artefaktais	38
3.12 pav. Diegimo modelis	39
4.1 pav. MergedObjects schema	40
4.2 pav. Papildyta Agents schema	40
4.3 pav. Vartotojo sąsaja pagrindiniame programos lange	41
4.4 pav. Įrankio klasių modelis	42
4.5 pav. Vartotojo sąsaja pagrindiniame <i>Protege</i> programos lange	43
4.6 pav. Individą nuskanti bendrinė žodžio forma.....	44

4.7 pav. Ontologijoje rasti ir apjungti panašūs individai	45
5.1 pav. SPARQL užklauso rezultatai ieškant individo „dalia“	47
5.2 pav. SPARQL užklauso rezultatai naudojant DISTINCT raktažodį	48
5.3 pav. SPARQL užklauso rezultatai ieškant individo „Obama“	49
5.4 pav. SPARQL užklauso rezultatai ieškant individo „obama“ naudojant DISTINCT raktažodį	50
5.5 pav. Pirmame straipsnyje paminėti ontologijos individai	51
5.6 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama	51
5.7 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Algirdas	52
5.8 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Hagelis	52
5.9 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Assadas.....	53
5.10 pav. Antrame straipsnyje paminėti ontologijos individai	53
5.11 pav. Antrame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama.....	53
5.12 pav. Trečiame straipsnyje paminėti ontologijos individai	54
5.13 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama	54
5.14 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Ericas	55
5.15 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Dalia	55
5.16 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Merkel	56
5.17 pav. Ketvirtame straipsnyje paminėti ontologijos individai	56
5.18 pav. Ketvirtame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama.....	57
5.19 pav. Ketvirtame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Linas	57
5.20 pav. Ketvirtame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Olekas	57
5.21 pav. Penktame straipsnyje paminėti ontologijos individai.....	58
5.22 pav. Penktame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama.....	58
5.23 pav. Penktame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Paulas	59
5.24 pav. Pirmo straipsnio lyginamoji lentelė.....	60
5.25 pav. Antro straipsnio lyginamoji lentelė	61
5.26 pav. Trečio straipsnio lyginamoji lentelė	61
5.27 pav. Ketvirto straipsnio lyginamoji lentelė	62
5.28 pav. Penkto straipsnio lyginamoji lentelė	62

TERMINŲ IR SANTRUMPŲ ŽODYNAS

Santrumpa, terminas	Paaškinimas
Ontologija	Tam tikros srities sąvokų visumos specifikuojimas
Ontologijos individai	Suprantami kaip objektai ir atpažįstami pagal vardus.
Panašūs individai	Jeigu individų bendrinė žodžio forma, kuri nurodoma <i>label_lemma</i> , yra vienoda, tada daroma prielaida, kad šie individai yra panašūs.
Label_lemma	Bendrinė žodžio forma, kuri nusako individo vardą
Egzemplioriai	Pagrindiniai žemo lygio ontologijų komponentai
Koncepcijos	Abstraktūs objektų rinkiniai
Atributai.	Naudojami specifinei objekto informacijai saugoti. Turi vardą ir reikšmę.
Anotavimas	Labai trumpai apibūdinamas publikacijos (knygos ar straipsnio) turinys.
Morfologija	1) žodžių formų ir jomis žymimų gramatinių reikšmių sistema; 2) mokslas apie žodžių reikšmines dalis, žodžių sandarą, klases, santykius. [1]
Žodžių kaityba	Tai žodžių kaitymas giminėmis, skaičiais, linksniais, nuosakomis, laikais, asmenimis, laipsniais ar kt., siekiant žodžių jungčių prasminio ir gramatinio taisyklingumo. [1]
PA	Panaudojimo atvejais
BPML	Verslo procesų modeliavimo kalba (angl. <i>Business Process Modeling Language</i>) yra verslo procesų ir juos palaikančių esybių abstraktus modelis
IS	Informacinė sistema
Label	Pavadinimas, kuris nurodo Individo pavadinimą
Comment	Komentarai, kuriais galima paašškinti arba detaliau aprašyti individą
Label_lt	Individo lietuviškas pavadinimas
ID	Identifikatorius
URL (angl. “ <i>Uniform Resource Locator</i> ”)	Informacinio resurso adresas internete.
V.	Vardininkas
K.	Kilmininkas
N	Naudininkas
G.	Galininkas

Santrumpa, terminas	Paaiškinimas
Įn.	Įnagininkas
Vt.	Vietininkas
Š.	Šauksmininkas
Vns.	Vienaskaita
Dgs.	Daugiskaita

ĮVADAS

Darbas parengtas pagal Informacinių sistemų inžinerijos magistrantūros studijų programos reikalavimus.

Darbo problematika ir aktualumas

Kiekvieną dieną internetiniuose puslapiuose, naujienų portaluose pateikiamas didžiulis informacijos kiekis dažnai yra perteklinis. Tokioje gausybėje informacijos labai lengva pasimesti ir nerasti reikalingos informacijos. Automatiškai analizuojant internetiniuose puslapiuose, naujienų portaluose publikuotus lietuviškus straipsnius susiduriama su problema, kad į ontologiją įrašoma pasikartojanti informacija. Semantiškai anotuotuose tekstuose rasti reikiamą informaciją būtų daug paprasčiau ir patogiau, jeigu ontologijoje būtų sudaryta galimybė automatiškai nustatyti panašius ontologijos individus ir juos susieti pagal panašumą.

Eksperimentui atlikti duota ontologija, kuri buvo sukurta Informacijos sistemų katedroje vykdant projektą „Lietuvių kalbos sintaksinės – semantinės analizės sistema tekstynui, lietuviškam internetui ir viešojo sektoriaus taikymams“, vykdomą pagal Ekonomikos augimo veiksmų programos 3 prioriteto „Informacinė visuomenė visiems“ įgyvendinimo priemonę Nr. VP2-3.1-IVPK-12-K „Lietuvių kalba informacinėje visuomenėje“. Ontologija užpildyta individualiais, kurie buvo sukurti automatiškai išanalizavus internete publikuotų lietuviškų straipsnių tekstus. Tokiu būdu kuriant ontologijos individus buvo sukurti individų dublikatai.

Darbo tikslas ir uždaviniai

Šio magistro darbo tikslas yra sudaryti galimybę automatiškai nustatyti panašius ontologijos individus, juos susieti ir tokiu būdu sumažinti informacijos ontologijoje dubliavimą.

Uždaviniai:

1. Išanalizuoti metodus, kurie taikomi teksto ir panašių individų suradimui.
2. Aprašyti individų suradimo taisykles, algoritmus.
3. Sukurti prototipą, kuris ieško panašių individų ir juos sujungia.
4. Eksperimentiškai ištirti panašių individų susiejimo teisingumą.
5. Apibendrinti tyrimo rezultatus.

Darbo rezultatai ir jų svarba

Sukurtas algoritmas ir parašytos SPARQL užklauskos sudaro galimybę automatiškai nustatyti panašius ontologijos individus ir juos susieti. Tokiu būdu sumažinamas informacijos ontologijoje dubliavimas. Apie egzempliorius sukauptas išanalizuotas informacijos rezultatas yra atvaizduojamas ontologijoje, kad vykdant paiešką kaskart nereikėtų analizuoti ontologijos individų panašumo.

Sukurtas sprendimas lengvai pritaikomas tokio tipo ontologijai, kurios struktūroje yra `label_lemma` (bendrinė žodžio forma), naudojant `label_lemma`'s bus sumažinti ontologijos individų pasikartojimai, sujungiant juos pagal panašumą ir taip siekiant pateikti konkrečią, struktūrizuotą informaciją.

Darbo struktūra

Darbe yra 7 skyriai. Pirmame skyriuje pateikta tyrimo objekto, srities ir problemos, naudotojų, esamų panašių ontologijos individų atpažinimo ir sujungimo galimybių analizė. Apibrėžtas darbo tikslas, uždaviniai ir rezultato kokybės kriterijai.

Antrame skyriuje pateikta panašių ontologijos individų atpažinimo ir sujungimo reikalavimų specifikacija, dalykinės srities modelis ir reikalavimų apibendrinimas.

Trečiame skyriuje pateiktas panašių ontologijos individų atpažinimo ir sujungimo algoritmas, sistemos architektūra, sistemos elgsenos ir realizacijos modeliai.

Ketvirtame skyriuje aprašyta sprendimo realizacija ir sukurto sprendimo testavimas.

Pentame skyriuje pateikiamas eksperimentas, kurio metu buvo parašytos SPARQL užklauskos, kurias įvykdžius išvedami panašūs ontologijos individai su nuorodomis į dokumentus.

Šeštame skyriuje pateikiamos galutinės darbo išvados.

Septintame skyriuje pateikiamas naudotos literatūros sąrašas.

1. PANAŠIŲ ONTOLOGIJOS INDIVIDŲ ATPAŽINIMO IR SUJUNGIMO GALIMYBIŲ TYRIMO PROBLEMINĖS SRITIES ANALIZĖ

Šiame skyriuje pateikiama su darbo problematika susijusios informacijos analizė, analizuojamas tyrimo objektas, esami problemos sprendimo būdai (technologijos, metodai), kitų autorių literatūra.

1.1. Analizės tikslas

Analizės tikslas – išnagrinėti panašių ontologijos individų suradimo ir sujungimo galimybes, geriau susipažinti su problema, nustatyti sprendimo eigą, atlikti esamų analogiškų sprendimų paiešką, egzistuojančius sprendimus išnagrinėti įvertinant jų privalumus ir trūkumus.

Analizei atlikti naudoti šie metodai:

- literatūros šaltinių studijavimas;
- analogiškų problemos sprendimų paieška bei analizė.

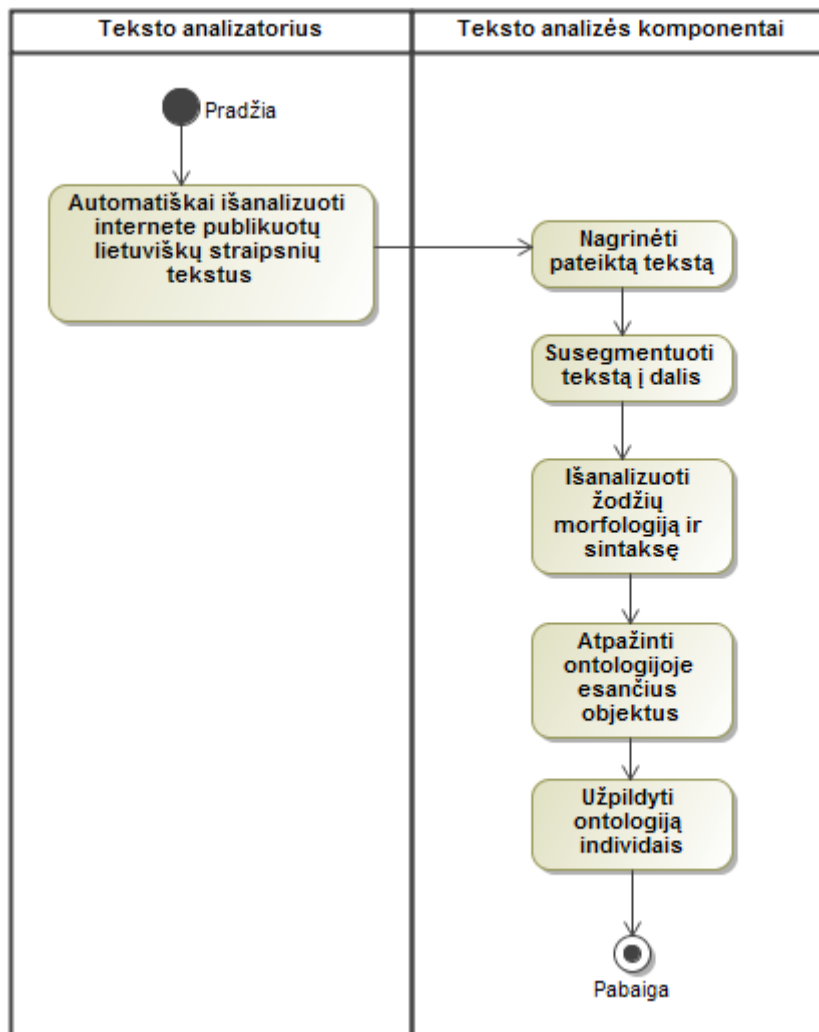
1.2. Tyrimo objektas, sritis ir problema

Magistrinio darbo tyrimo objektas – gauta individualiai užpildyta ontologija.

Šio magistrinio darbo tyrimo sritis – panašių ontologijos individų suradimas ir sujungimas.

Kiekvieną dieną internetiniuose puslapiuose, naujienų portaluose pateikiamas didžiulis informacijos kiekis dažnai yra perteklinis. Tokioje gausybėje informacijos labai lengva pasimesti ir nerasti reikalingos informacijos. Automatiškai analizuojant internetiniuose puslapiuose, naujienų portaluose publikuotus lietuviškus straipsnius susiduriama su problema, kad į ontologiją įrašoma pasikartojanti informacija. Semantiškai anotuotuose tekstuose rasti reikiamą informaciją būtų daug paprasčiau ir patogiau, jeigu ontologijoje būtų sudaryta galimybė automatiškai nustatyti panašius ontologijos individus ir juos susieti pagal panašumą.

Eksperimentui atlikti duota ontologija, kuri buvo sukurta Informacijos sistemų katedroje vykdant projektą „Lietuvių kalbos sintaksinės – semantinės analizės sistema tekstynui, lietuviškam internetui ir viešojo sektoriaus taikymams“, vykdomą pagal Ekonomikos augimo veiksmų programos 3 prioriteto „Informacinė visuomenė visiems“ įgyvendinimo priemonę Nr. VP2-3.1-IVPK-12-K „Lietuvių kalba informacinėje visuomenėje“. Ontologija užpildyta individualiais, kurie buvo sukurti automatiškai išanalizavus internete publikuotų lietuviškų straipsnių tekstus. Išanalizuoti straipsnių tekstai buvo paduodami eilei teksto analizės komponentų, kurie nagrinėjo pateiktą tekstą, jį susegmentavo į dalis, išanalizavo žodžių morfologiją, kiek įmanoma – sintaksę ir pagal įvairias taisykles atpažino duotoje ontologijoje esančius objektus – asmenis, vietas, organizacijas ir t.t. (Šio proceso veiklos diagrama pateikta 1.1 pav.) Tokiu būdu kuriant ontologijos individus buvo sukurti individų dublikatai.



1.1 pav. Ontologijos užpildymo individais veiklos procesas

Sistemoje, kuri tokiu būdu kuria duomenis, esantys objektai gali būti paminėti kelis kartus, tačiau ne visą laiką algoritmas gali pažinti, kad, pavyzdžiui, tas pats kelis kartus paminėtas asmuo tikrai yra vienas ir tas pats. Asmuo skirtinguose sakiniuose gali būti užrašytas pilnu vardu (pvz., Dalia Grybauskaitė) arba sutrumpintu (pvz., D. Grybauskaitė), taip pat gali būti paminėtas skirtinguose linkniuose (pvz., Dalios Grybauskaitės, D. Grybauskaitę ir pan.). Programų algoritmai tą patį asmenį gali atpažinti kaip kelis skirtingus egzempliorius.

1.2 pav. pateikti teksto fragmentai su pasikartojančiais egzemplioriais. Šiuos straipsnius automatiškai išanalizavus, kiekvieną kartą tekste radus paminėtą asmenį, būtų kuriamas naujas individas. Algoritmas šiuose dviejuose skirtinguose tekstuose esančius objektus anotuos, sukurs kelis skirtingus ontologijos klases Asmuo individus ir atpažins kaip skirtingus egzempliorius, nors iš teksto konteksto suprantama, kad tai yra vienas ir tas pats asmuo.

Įsilaužėliai KGB veiklą viešinančioje interneto svetainėje paskelbė šmeižikišką informaciją, esą prezidentė **Dalia Grybauskaitė** dirbo KGB, pirmadienį pranešė svetainę administruojantis Lietuvos gyventojų genocido ir rezistencijos tyrimo centras.

D. Grybauskaitės duomenys ir pravadė „Magnolija“ pridėta prie KGB agentų sąrašo.

Lietuvos gyventojų genocido ir rezistencijos tyrimo centro direktorė **Teresė Birutė Burauskaitė** pareiškė, kad ši informacija „neatitinka tikrovės ir klaidina visuomenę“.

„Mums pranešė žurnalistas apie ketvirtą popiet. Dabar mūsų programuotojai dirba. **D. Grybauskaitė** įtraukta į agentų sąrašą, kurį mes skelbiame“, - **BNS** sakė **T. B. Burauskaitė**.

Prezidentūra tai vertina kaip eilinę informacinę provokaciją, teigiama Prezidentūros DELFI atsiųstame komentare.

Interneto svetainėje www.kgbveikla.lt įsilaužėliai taip pat patalpino archyvinius dokumentus primenančią medžiagą, susijusią su **D. Grybauskaite**. Kaip nurodoma, ten pateikiamas tuometinio Leningrado Ždanovo universiteto diplomas, jo priedas, ranka rašyta tariamo **D. Grybauskaitės** vidurinės baigimo atestato kopija, kiti jos studijomis susiję dokumentai.

„Byloje“ taip pat patalpintos **D. Grybauskaitės** studijų užrašų fotokopijos, charakteristika iš darbovietės, sveikatos kortelės duomenys.

KGB veiklą tyręs istorikas ir Seimo narys **Arvydas Anušauskas** **BNS** sakė, kad interneto svetainėje patalpinta „elementari studentiška byla“.

„Studentiškos bylos paprastai saugomos universitetuose“, - teigė jis. Atsižvelgiant į tai, kad medžiaga patalpinta įsilaužus į interneto svetainę, **A. Anušauskas** teigė neabejojantis, kad tai yra Rusijos Federalinės saugumo tarnybos FSB darbas.

Aprašyme daryta lietuvių kalba teigiama, kad „tai yra studentės **D. Grybauskaitės** asmens bylos fotokopijos jos mokymo sovietmečio laikų Leningrade metu“. Anonimas teigia atstovaujantis nežinomą organizaciją „Kumštis“.

Apie šį incidentą pranešta praėjus kelioms valandoms po to, kai Lietuvos žvalgybos institucijos paskelbė ataskaitas apie Rusijos keliamas grėsmes.

D. Grybauskaitė 1983 metis baigė politinę ekonomiją tuometiniame Leningrado universitete, studijų metais dirbo kailių fabrike.

1988 metais Maskvos visuomenės mokslų akademijoje ji apgynė disertaciją ir įgijo ekonomikos mokslų kandidato laipsnį.

Prezidentė **Dalia Grybauskaitė** siūlo kuo skubiau užbaigti iliustracijos procesą, paviešinant visą archyvuose turimą medžiagą, susijusią su KGB veiksmis įtraukiant žmones į savo veiklą, kartu paaiškinant dokumentų prasmę. Taip ji pareiškė trečiadienį susitikusi su Seimo Nacionalinio saugumo ir gynybos komiteto (NSGK) nariais. „Tai yra vienintelis aiškus, greitas, suprantamas ir būtinas kelias pabaigti du dešimtmečius besitęsiančią ir niekaip nesibaigiančią iliustracijos temą“, - šalies vadovės žodžius citavo jos atstovas spaudai Linas Balsys.

Savo ruožtu NSGK pirmininkas konservatorius-krikdemas **Arvydas Anušauskas** teigė manantis, kad teisinę iliustraciją reikia pratęsti bent iki 2012 m. ieškant apie ryšius su KGB nepripažinusių asmenų.

Nori švietimo apie SSRS spec. tarnybų veiklą

Pasak **L. Balsio**, anksčiau jau buvo suteikta galimybė prisipažinti žmonėms, įtrauktiems į Sovietų Sąjungos specialiųjų tarnybų veiklą. Apie šių asmenų buvusią veiklą žinoma, jiems buvo garantuota valstybės apsauga. Tačiau procesas iki šiol nebaigtas: Iliustracijos komisija dirba, esama bylų, bet didžioji dalis jų teismuose baigiasi niekuo, nes dėl medžiagos trūkumo teismai negali priimti sprendimo.

„Iliustraciją reikia kuo skubiau užbaigti, paviešinant visą archyvuose turimą medžiagą, susijusią su tuo, kaip KGB ir kitos specialiosios tarnybos įtraukdavo Lietuvos gyventojus į savo veiklą, tos veiklos metodus, - prezidentę citavo **L. Balsys**. - Informaciją reikia padaryti viešą, dokumentus skelbti (techninės detalės, kaip tai padaryti).“

D. Grybauskaitės manymu, kartu su dokumentais reikia pateikti kvalifikuotą, profesionalų paaiškinimą, kad būtų lengviau suprasti, ką tam tikras dokumentas reiškia ir kokią įtaką jis galėjo turėti žmogaus veiklai bei likimui, ar šis asmuo galėjo būti užverbuotas ir tapti agentu. „Tai turėtų būti masiškas, aiškus žmonių švietimas apie spec. tarnybų veiklą ir jų padarytą žalą“, - apibendrinio **L. Balsys**.

Apie archyvinės medžiagos viešinimą prezidentė kalbėjo jau pernai. **A. Anušauskas** yra užregistravęs vadinamojo Iliustracijos įstatymo pataisas, kuriose numatytas dokumentų viešinimas ir apibrėžta, kaip tai turi vykti. NSGK nariai artimiausiu metu žada pateikti savo išvadą dėl pataisų.

Dalis bylų buvo išvežta ar sunaikinta

„Galiu pasakyti savo nuomonę - pereinamasis laikotarpis prieš nutraukiant teisinę iliustraciją turi būti, bent iki 2012 m. (...) Iš patirties žinau, kad yra nepripažinusių ar atlikusių „kabinetinę iliustraciją“ - ne Aukščiausiosios Tarybos komisijai KGB veiklai tirti, ne pagal įstatymus, o tiesiog ateidavo į buvusių Valstybės saugumo vadovų kabinetus ir sutardavo dėl kažkokių mums nežinomų dalykų“, - po susitikimo su prezidente kalbėjo **A. Anušauskas**.

Pasak jo, dokumentų viešinimas nereikš žmonių plūdimo į archyvus: represuotųjų artimieji, istorikai esą ir dabar gali sklaidyti dokumentus, neleidžiama susipažinti tik su duomenimis, kurie yra susiję su jau prisipažinusiais asmenimis. Ar žmogus bendradarbiavo su sovietų spec. tarnybomis, turėtų įvertinti teismas.

Iki šiol apie bendradarbiavimą su KGB yra prisipažinę apie 1,5 tūkst. žmonių, kai kurie atliko „kabinetinę iliustraciją“. Maždaug 5 tūkst. bylų pirmaisiais Nepriklausomybės metais buvo išvežta į Maskvą. Dar apie 105 tūkst. agentų bylų buvo sunaikinta 1989-1990 m. Tačiau, kaip pažymėjo istoriko išsilavinimą turintis **A. Anušauskas**, šios bylos apėmė visą KGB gyvavimo laikotarpį, įskaitant net pokario metus. Vykstant partizaninei kovai esą daug žmonių buvo formaliai įrašomi į agentus - dėl skaičiaus. Tuomet jų būta 27 tūkst., įskaitant esą užverbuotus miškininkus, medkircius ir pan. Dalis žmonių, kurie turėjo ryšių su KGB, emigravo ar mirė.

1.2 pav. Teksto fragmentai (straipsniai iš internetinio portalo www.delfi.lt) su pasikartojančiais egzemplioriais

Analizuojamoje ontologijoje buvo rastas pavyzdys - asmuo Žanas Klodas ir asmuo Van Damas. Dabartiniai komponentai, kurie analizuoja lietuvišką tekstą, pagal tam tikras taisykles nustatė, kad tai yra du skirtingi egzemplioriai, nors realiame pasaulyje Žanas Klodas Van Damas yra vienas ir tas pats asmuo. Pagal teksto analizės algoritmo taisykles, jeigu yra du žodžiai iš dviejų didžiųjų raidžių, tai greičiausiai bus vardas ir pavardė. Šios taisyklės nepritaikytos kelis vardus turintiems asmenims.

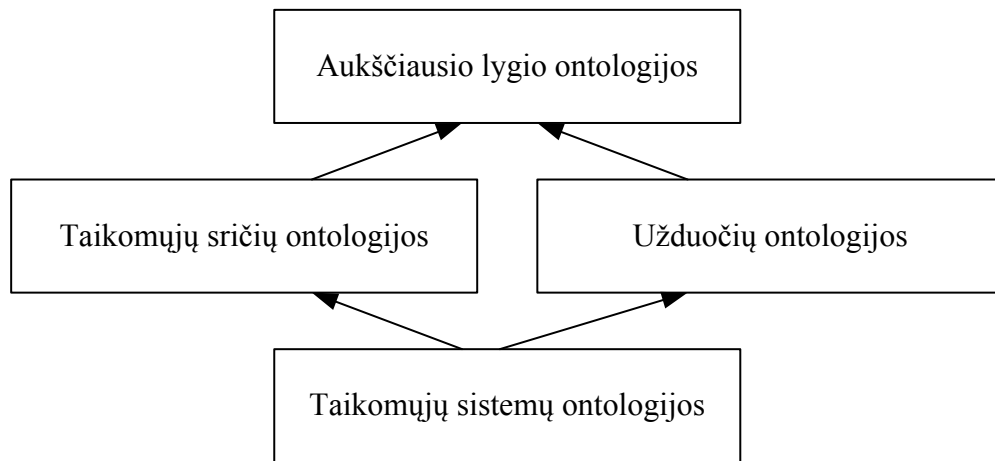
Nekokybiški teksto analizės rezultatai dažnai yra semantinio anotavimo komponento taisyklių ir leksinės analizės komponentų ribotumas, jų netobulumas.

1.3. Ontologijos ir jos individų analizė

Ontologija - tai tam tikros srities sąvokų visumos specifikuojamas išreikštu pavidalu (*“Explicit specification of a conceptualization”*, T.R.Gruber, 1993 m.). Ontologija apibrėžia nagrinėjamos srities esybių tipus, jų tarpusavio sąryšius, galiojančias taisykles, dėsningumą ir aksiomas [2]. Dažniausiai nagrinėjamos tik tam tikros srities ontologijos, nes išnagrinėti viso pasaulio ontologiją, suvokti jos struktūrą būtų be galo sudėtinga. Kompiuterijos ir dirbtinio intelekto srityse ontologijos terminas pradėtas naudoti nuo 1990 metų.

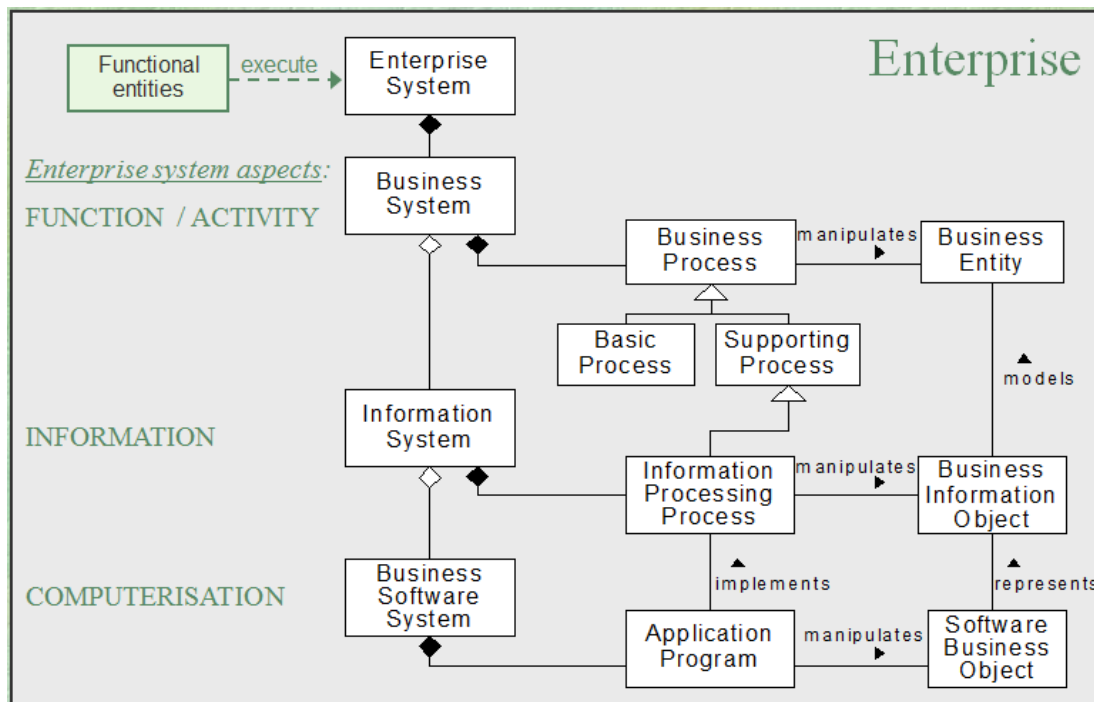
N. Guarino požiūriu, ontologijos yra ypatingos rūšies žinių bazės, aprašančios faktus, kuriuos tam tikra naudotojų grupė laiko visada teisingais, remiantis sutartomis naudojamu žodyno terminų prasmėmis [3].

N. Guarino straipsnyje [3] išskiriamos keturios ontologijų rūšys (žr. 1.3 pav.).



1.3 pav. Svarbiausios ontologijų rūšys [3]

Pirmoji ontologijų rūšis – tai aukščiausio lygio ontologijos. Jos aprašo bendras sąvokas, kurios nepriklausomas nuo konkrečios problemos. [3] Antroji - taikomųjų sričių ontologijos. Šios ontologijos aprašo konkrečių sričių žodynus (pvz., informatikos). Pavyzdys - M. Uschold'o įmonių ontologija [4]. Trečioji ontologijų rūšis – tai užduočių ontologijos, kurios aprašo konkrečių užduočių žodynus. Pavyzdys - verslo procesų modeliavimo kalba (BPML [5]), kuri apima visus įmonių verslo procesus, apibrėžia abstrakčių ir vykdomų procesų formalų modelį. Na ir ketvirtoji ontologijų rūšis, paminėta N. Guarino straipsnyje – tai taikomųjų sistemų ontologijos, kurios aprašo taikomųjų sričių, užduočių ontologijų specializacijos konceptus. [3] Pavyzdys - Unifikuota įmonių inžinerijos ontologija [6] (žr. 1.4 pav. *Įmonių inžinerijos ontologija*), kurioje apibrėžiami verslo sistemų, informacinių ir programinių sistemų tarpusavio priklausomybės.



1.4 pav. Įmonių inžinerijos ontologija [6]

Pagal paskirtį ontologijos skirstomos į [7] žinių vaizdavimo, lingvistines, užduočių, nagrinėjimo sričių, metodų, bendrąsias, taikomųjų programų, visuotinai naudojamų sąvokų, aukščiausio lygio ir meta ontologijas.

Ontologijų elementai – egzemplioriai (individai), konceptai (klasės), atributai ir ryšiai [8].

- Ontologijos individai suprantami kaip objektai ir atpažįstami pagal vardus. Ontologijose individo vardas yra prasmingas pavadinimas. Individo pavadinimas dažniausiai sugeneruojamas automatiškai.
- Konceptai gali susidėti iš egzempliorių, kitų klasių arba egzempliorių ir klasių junginių.
- Atributai naudojami specifinei objekto informacijai saugoti.

Žemiau pateikta gautos ontologijos schemos dalis, kuri aprašo šiame darbe nagrinėjamus ontologijos individus. Nuspręsta nagrinėti tik Asmenų klasės individus, nes konkrečių asmenų paieška ir su jais susijusi informacija šiuo metu yra ko gero labiausiai naudojama įvairiose veiklos srityse. Labai panaši situacija būtų nagrinėjant vietovių arba organizacijos pavadinimų objektus, tiesiog būtų taikomos kitos taisyklės ontologijos individų susiejimui.

```

<!-- http://semantika.lt/ns/Agents#agent -->

<owl:Class rdf:about="&agents;agent">

  <rdfs:label xml:lang="lt">agentas</rdfs:label>

  <rdfs:subClassOf rdf:resource="&sem;object"/>

  <ann:label_lt xml:lang="lt">agentas</ann:label_lt>

  <ann:label_sbvr xml:lang="lt">agentas</ann:label_sbvr>

</owl:Class>

<!-- http://semantika.lt/ns/Agents#person -->

```

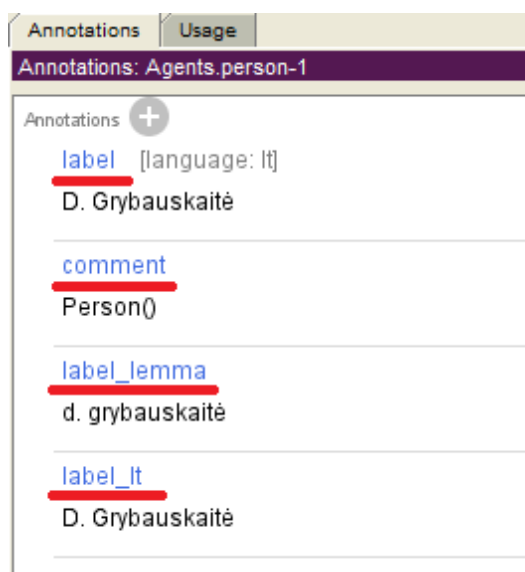
```

<owl:Class rdf:about="&agents;person">
  <rdfs:label xml:lang="lt">asmuo</rdfs:label>
  <rdfs:subClassOf rdf:resource="&agents;agent"/>
  <ann:label_sbvr xml:lang="lt">asmuo</ann:label_sbvr>
  <ann:label_lt xml:lang="lt">asmuo</ann:label_lt>
</owl:Class>

```

Visi individai nagrinėjamoje ontologijos struktūroje (žr. 1.5 pav. *Individai ontologijų struktūrose*) aprašomi anotacijomis:

- label (pavadinimas, kuris nurodo Individo pavadinimą);
- comment (komentarai, kuriais galima paaiškinti arba detaliau aprašyti individą);
- label_lemma (bendrinė žodžio forma);
- label_lt (lietuviškas pavadinimas).



1.5 pav. Individų ontologijos struktūroje aprašymas

Analizuojamoje ontologijoje kiekvienas atpažintas egzempliorius turi nuorodą į dokumentą. Pavyzdžiui, jeigu asmuo Jonas Jonaitis yra paminėtas viename dokumente, tai jis turės nuorodą į tą vieną dokumentą. Jeigu kitame dokumente bus minimas Jonas Jonaitis, tuomet jis turės kitą identifikatorių ir nuorodą į kitą dokumentą. Jeigu skirtinguose dokumentuose yra paminėti du Jonai Jonaičiai, tai ar jie vienodi ar ne, niekas nežinos. Galima daryti prielaidą, kad galbūt šie du asmenys yra vienas ir tas pats asmuo.

Žemiau pateikiamas analizuojamos ontologijos RDF failo fragmentas, kuriame matomi pasikartojantys ontologijos individai „Vytautui Gapšiui“. Šie ontologijos individai skirtingi, nes turi skirtingus identifikatorius.

```

<http://semantika.lt/ns/Agents#person~0163d8b6-72f3-4eea-bc89-035d1fcf15dd~Agents.person-3> rdfs:comment "Person (Vytautui Gapšiui)" .

```

```

<http://semantika.lt/ns/Agents#person~4f93357d-50e6-4087-baef-e8fe6e137536~Agents.person-3> rdfs:comment "Person (Vytautui Gapšiui)" .

```

```

<http://semantika.lt/ns/Agents#person~22be454c-9e7d-436c-9f6f-766932c4efce~Agents.person-3> rdfs:comment "Person (Vytautui Gapšiui)" .

```

<http://semantika.lt/ns/Agents#person~59027521-c242-4874-bf9e-6856bdfb6f3c~Agents.person-3> rdfs:comment "Person(Vytautui Gapšiui)" .

<http://semantika.lt/ns/Agents#person~2df229a5-7d48-4918-bc92-6737c2578ca6~Agents.person-4> rdfs:comment "Person(Vytautui Gapšiui)" .

<http://semantika.lt/ns/Agents#person~3710406f-dbaa-42f7-9200-af2d990baea9~Agents.person-4> rdfs:comment "Person(Vytautui Gapšiui)" .

<http://semantika.lt/ns/Agents#person~231a9f3e-16b9-4fee-8415-a0c230f86ab8~Agents.person-4> rdfs:comment "Person(Vytautui Gapšiui)" .

Nagrinėjamoje ontologijoje individo prasmingas pavadinimas dedamas į anotacijas (label, label_lemma). Label_lemma anotaciją buvo nuspręsta naudoti tik Informacijos sistemų katedroje sukurtoje ontologijoje vykdančią projektą „Lietuvių kalbos sintaksinės – semantinės analizės sistema tekstynui, lietuviškam internetui ir viešojo sektoriaus taikymams“, vykdomą pagal Ekonomikos augimo veiksmų programos 3 prioriteto „Informacinė visuomenė visiems“ įgyvendinimo priemonę Nr. VP2-3.1-IVPK-12-K „Lietuvių kalba informacinėje visuomenėje“, taip siekiant apeiti lietuvių kalbos žodžių kaitymo sukeltas problemas. Kitose ontologijose, kuriose taip pat yra asmenų klasės, tokios anotacijos gali ir nebūti.

Label anotacijoje nurodoma individo vardo ar pavadinimo reikšmė. Jeigu individų bendrinė žodžio forma, kuri nurodoma label_lemma anotacijoje, yra vienoda, tada daroma prielaida, kad šie individai yra panašūs. Pavyzdžiui, turime individą *Andriui Kupčinskui* ir individą *A. Kupčinsko*, šių individų abi label_lemma reikšmės yra *Andrius Kupčinskas*. Taigi, daroma prielaida, kad šie individai yra panašūs.

Teoriškai reikėtų vertinti ir kitas žinomas asmens savybes, ne tik vardą, pavardę, bet ir kur asmuo dirba, kur gyvena, ką mėgsta ir pan. Tokiu atveju turint du asmenis tokiais pat vardais ir pavardėmis, ir žinant papildomas tų asmenų savybes, panašumas tik labiau didėtų.

1.4. Tyrimo objekto naudotojų analizė

Analizuojamos ontologijos naudotojai yra sistemos kūrėjai, kurie tvarko, ruošia informaciją. Sistemos kūrėjai nori sudaryti galimybę naudotojams greičiau rasti juos dominančius asmenis, organizacijas, vietas, kurias galima apjungti kaip panašias. Jie sprendžia, kaip paruošti duomenis susietai, sutvarkyti informacijai gauti.

1.5. Panašių ontologijos individų atpažinimo ir sujungimo galimybių analizė

Šiame skyriuje apžvelgti metodai (Levenshtein'o atstumas, Damerau–Levenshtein'o atstumas ir Hamming'o atstumas), kurie galėtų būti panaudoti nustatant, kurie individai tarpusavyje panašūs, tam kad būtų galima surasti konkrečią informaciją apie asmenis, vietas ar organizacijų pavadinimus.

Hamming'o atstumas tarp dviejų vieno ilgio eilučių yra skaičius pozicijų, per kurias atitinkami simboliai yra skirtingi. Kitaip pasakius, tai yra minimalus keitimų skaičius, norint pakeisti vieną eilutę į kitą, arba minimalus klaidų skaičius, dėl kurių viena eilutė galėjo tapti kita. [9]

Hamming'o atstumo pavyzdys:

- 6531391 ir 6561691 yra 2.
- 2143896 ir 2233496 yra 3.
- „karolina“ ir „katerina“ yra 3.

Levenshtein'o atstumas tarp dviejų žodžių yra minimalus redagavimų po vieną simbolį skaičius (pvz., įterpimų, ištrynimų arba pakeitimų) reikalingų vieno žodžio pakeitimui į kitą žodį. [10] Šis atstumas pavadintas Vladimir'o Levenshtein'o garbei, kuris sugalvojo šį metodą 1965 metais. [11]

Damerau–Levenshtein‘o atstumas - tai atstumo matavimo vienetas, gautas įtraukiant gretimų simbolių perkėlimą į atstumą tarp dviejų eilučių [12].

Savo konstruktyviuose užrašuose [13] Damerau ne tik išskyrė keturias koregavimo operacijas, bet taip pat pareiškė, kad jos atitinka daugiau nei 80% visų žmogiškų rašybos klaidų. Damerau's užrašuose laikomos tik tos rašybos klaidos, kurios gali būti ištaisytos su viena koregavimo operacija.

Damerau–Levenshtein‘o atstumu siekiama išmatuoti atstumą tarp žmogaus rašybos klaidų, kad būtų galima pagerinti tokias programas, kaip rašybos klaidų tikrinimas [14].

Damerau–Levenshtein‘o atstumas skiriasi nuo klasikinio Levenshtein‘o atstumo didesniu leistinų operacijų kiekiu. Klasikinis Levenshtein‘o atstumas leidžia tik įterpimo, ištrynimo ir pakeitimo operacijas [15].

Levenshtein‘o atstumo ir Damerau–Levenshtein‘o atstumo pavyzdys:

- bitė → ritė ("b" pasikeičia į "r")
- vagis → vagių ("s" pasikeičia į "ų").

Damerau–Levenshtein‘o atstumo algoritmas paieškos rezultatuose išvestų ganėtinai panašius individus į ieškomą, jei rašant buvo suklysta viena – dviem raidėmis. Pavyzdžiui, žurnalistas rašydamas pavardę *D. Grybauskaitės* suklydo ir parašė *D. Grybausakitės* (žr. 1.6 pav.). Šiuo atveju panašių ontologijos individų sujungimo įrankis pagal Damerau–Levenshtein‘o atstumo algoritmą atrinktų individą *D. Grybausakitės* kaip panašų į individą *D. Grybauskaitė*.

Interneto svetainėje www.kgbveikla.lt įsilaužėliai taip pat patalpino archyvinius dokumentus primenančią medžiagą, susijusią su D. Grybauskaite. Kaip nurodoma, ten pateikiamas tuometinio Leningrado Ždanovo universiteto diplomas, jo priedas, ranka rašyta tariamo D. Grybauskaitės vidurinės baigimo atestato kopija, kiti jos studijomis susiję dokumentai.

„Byloje“ taip pat patalpintos D. Grybausakitės studijų užrašų fotokopijos, charakteristika iš darbovietės, sveikatos kortelės duomenys.

KGB veiklą tyręs istorikas ir Seimo narys Arvydas Anušauskas BNS sakė, kad interneto svetainėje patalpinta „elementari studentiška byla“.

„Studentiškos bylos paprastai saugomos universitetuose“, - teigė jis. Atsižvelgiant į tai, kad medžiaga patalpinta įsilaužus į interneto svetainę, A. Anušauskas teigė neabejojantis, kad tai yra Rusijos Federalinės saugumo tarnybos FSB darbas.

Aprašyme daryta lietuvių kalba teigiama, kad „tai yra studentės D. Grybauskaitės asmens bylos fotokopijos jos mokymo sovietmečio laikų Leningrade metų“. Anonimas teigia atstovaujantis nežinomą organizaciją „Kumštis“.

1.6 pav. Teksto fragmentas (straipsnis iš internetinio portalo www.delfi.lt) su klaidingai parašyta asmens pavarde

Visgi, lyginant skirtingo ilgio teksto eilutes, arba eilutes, kuriose gali būti tikėtini pridėjimai, atėmimai ir įterpimai, pažangesnė metrika yra Levenshtein‘o atstumas. Levenshtein‘o atstumas taip pat gali būti suskaičiuotas tarp dviejų ilgų eilučių, bet šio skaičiavimo kaina, kuri yra apytiksliai proporcinga dviejų eilučių ilgiui, padaro šį skaičiavimą nepraktišku [16].

Magistrinio darbo tikslui pasiekti Hamming‘o atstumas nėra tinkamas, nes jis turi apribojimą, kad lyginamos eilutės turi būti vienodo ilgio. Tikslui pasiekti galėtų būti naudojamas Damerau–

Levenshtein'o atstumas, kuris leidžia dviejų gretimų ženklų įterpimą, ištrynimą, koregavimą ir perkėlimą.

1.5.1. Žodžių morfologinės savybės

Siekiant sujungti panašius ontologijos individus taip pat labai svarbu įvertinti ir žodžių morfologines savybes.

Analizuojamoje ontologijoje visi ontologijos individų pavadinimai pateikti lietuvių kalba. Šioje ontologijoje individų pavadinimai yra sulemuoti – visi žodžiai atvesti į antraštinę formą (jei kalbant apie daiktavardžius, tai į vienaskaitos vardininko linksnį). Panašumas nustatinėjamas pagal ontologijos individo Label_lemma reikšmę. Label_lemma naudojama, kai žodis yra žinomas morfologinėje duomenų bazėje ir kai teksto analizės komponentai moka pagaminti tą lemą. Tačiau nevisi lietuviški vardai ir pavardės yra toje morfologinėje duomenų bazėje, nevisi jie gali būti teisingai sulemuoti. Jeigu žodis nebuvo atpažintas dėl gramatinių klaidų, arba tai yra naujas žodis ir jo nebus morfologinėje duomenų bazėje, tuomet tas žodis Label_lemma reikšmėje bus įrašytas toks, kokį rado teksto analizatoriai (bet kuriame linksnyje (kilmininko, galininko ir t.t), formoje (daugiskaitoje, vienaskaitoje), moteriškoje ar vyriškoje giminėje ir pan.). Pavyzdžiui, automatiškai išanalizavus internete publikuotą lietuvišką straipsnio tekstą buvo rastas asmuo Jeronimui Miliauskiui. Šis asmuo nebuvo žinomas morfologinėje duomenų bazėje, dėl to šio asmens Label ir Label_lemma reikšmėse bus įrašyta Jeronimui Miliauskiui.

Dalis asmenų vardų ir pavardžių yra nelietuviški, sulietuvinti (pavyzdžiui, asmuo Barack Obama gali būti užrašytas įvairiai, kaip Barakas Obama, Barak Obama ir t.t.). Tokių žodžių gali nebūti morfologinėje duomenų bazėje, dėl to jie niekada nebus sulemuoti ir bus pateikiami įvairiausiose formose, linksniuose. Vien tik lemų naudojimas neduoda garantijos, kad visos lemos visais atvejais bus korektiškos ir kad tai tikrai žodžio antraštinė forma.

Skirtingos situacijos ir skirtinga duomenų kokybė bus ieškant asmenų, organizacijų pavadinimų, vietovių ir pan. Situacijos ieškant panašių ontologijos individų skiriasi tam tikrais taisyklių rinkiniais. Taisyklės priklauso nuo to, kaip programa tekstą skaldo į dalis, sakinius, žodžius. Taisyklės algoritmuose ne visada gerai suveikia. Jeigu taip atsitinka, tada ir morfologiniai analizatoriai blogai išanalizuoja tekstą.

Kaitant žodžio forma keičiasi, nors pagal prasmę jis yra tas pats. Tikriniai daiktavardžiai, išreiškiantys asmenų vardus, yra kaitomi visuose linksniuose, priklausomai nuo to, kokiame vaidmenyje parašyti tekste. Jeigu asmuo, paminėtas vienaskaitos vardininko linksnyje, dalyvaus kitame vaidmenyje, tai jo vardo užrašymo forma tekste skirsis galūne (bus pakeistas linksnis). Pavyzdžiui, „Kauno ponija **Livija Levienė** dalyvavo gėlių parodoje. **Livijai Levienei** labai patiko baltos rožės.“

Nelietuviškų tikrinių vardų rašymo taisyklės nuolat keičiasi, todėl kaskart reikėtų pasidomėti, ką šiuo klausimu rašo VLKK (Valstybinė lietuvių kalbos komisija, <http://www.vlkk.lt/>).

Lietuviškas vardas ir pavardė **Petras Kazlauskas** linksniuojamas:

	Vns.	Dgs.
V.	Petras Kazlauskas	Petrai Kazlauskai

	Vns.	Dgs.
K.	Petro Kazlauskio	Petrų Kazlauskų
N	Petrui Kazlauskui	Petrams Kazlauskams
G.	Petrą Kazlauską	Petrus Kazlauskus
Įn.	Petru Kazlausku	Petrais Kazlauskais
Vt.	Petre Kazlauske	Petruose Kazlauskuose
Š.	Petrai Kazlauskai	Petrai Kazlauskai

Gautoje ontologijoje daugiskaitos formoje tikrinių daiktavardžių nėra, nes asmuo, pavadintas, pvz., Petrais, nebus atpažintas kaip konkretus asmuo. Petras Kazlauskas – tai jau konkretus asmuo ir jis tampa asmenų klasės egzemplioriumi.

1.6. Darbo tikslas, uždaviniai ir siekiami privalumai

Šis poskyris apibrėžia darbo tikslą ir numatomus uždavinius, kurie bus sprendžiami, siekiant išspręsti kilusią problemą.

Darbo tikslas – sudaryti galimybę automatiškai nustatyti panašius ontologijos individus, juos susieti ir tokiu būdu sumažinti informacijos ontologijoje dubliavimą.

Magistro baigiamajame darbe suformuoti uždaviniai:

1. Išanalizuoti metodus, kurie taikomi teksto ir panašių individų suradimui.
2. Aprašyti individų suradimo taisykles, algoritmus.
3. Sukurti prototipą, kuris ieško panašių individų ir juos sujungia.
4. Eksperimentiškai ištirti panašių ontologijos individų susiejimo teisingumą.
5. Apibendrinti tyrimo rezultatus.

Žemiau pateikiami kokybės kriterijai, pagal kuriuos bus vertinamas mano sukurto algoritmo gerumas.

- Teisingai nustatyti panašumai (truePositives): nustatytų teisingų panašumų skaičius ontologijoje.
- Klaidingai nustatyti panašumai (falsePositives): klaidingų elementų, palaikytų teisingais (panašiais), skaičius.
- Nesurasti panašumai (falseNegatives): praleistų elementų (klaidingai nepalaikytų panašiais) skaičius ontologijoje.
- Tikslumas: teisingai nustatytų panašumų santykis su visais (teisingais ir neteisingais) nustatytais panašumais

$(\text{truePositives} / (\text{truePositives} + \text{falsePositives}))$ [17].

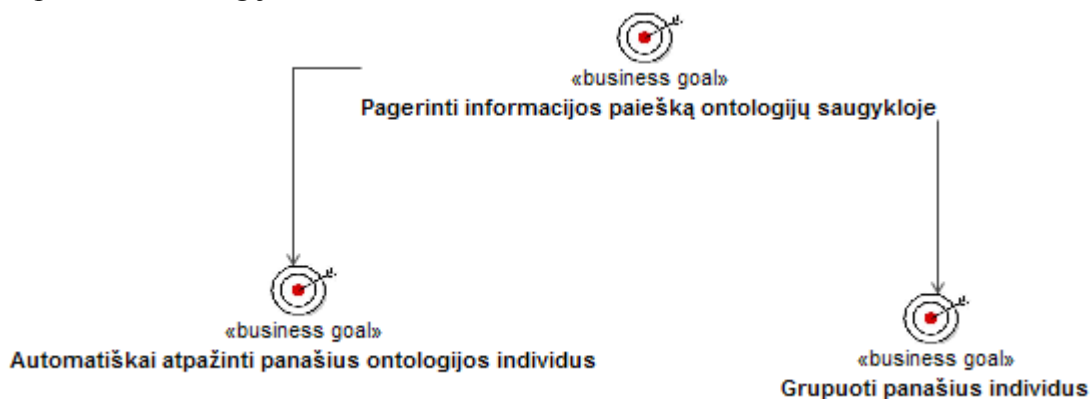
- Atkūrimas: teisingai nustatytų panašumų santykis su visais (nustatytais ir nenustatytais) panašumais
 $(\text{truePositives} / (\text{truePositives} + \text{falseNegatives}))$ [17].
- Kokybė: harmonizuota tikslumo ir atkūrimo išraiška
 $(2 * (\text{tikslumas} * \text{atgaminimas}) / (\text{tikslumas} + \text{atgaminimas}))$.

Numatomas algoritmo praktinis taikymas pagerins paieškos rezultatus grupuojant besidubliuojančius, panašius įrašus.

1.7. Siekiamo sprendimo apibrėžimas

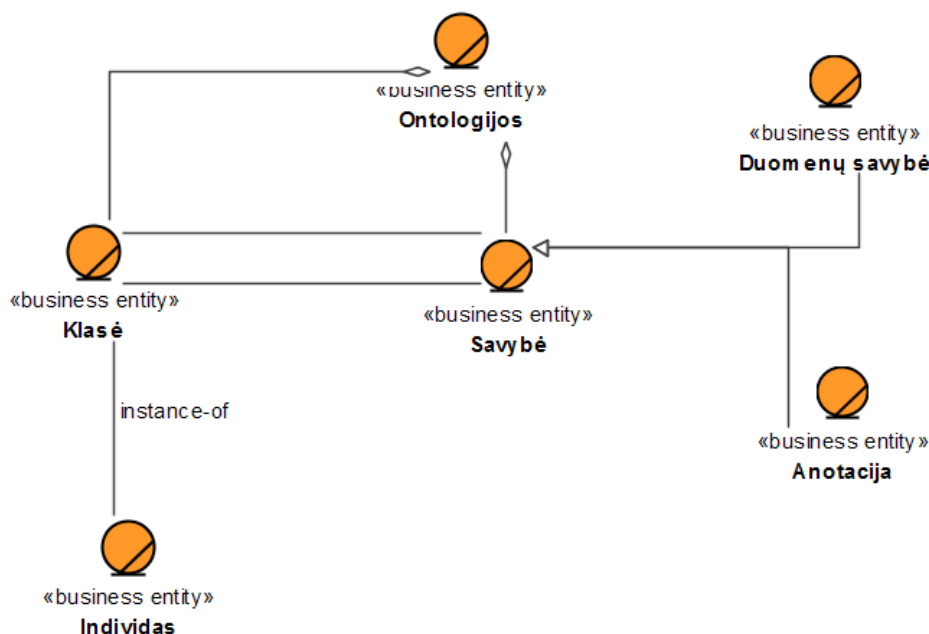
Siekiamas sprendimas – sukurti algoritmą, kuris susies rastus panašius ontologijos individus.

Tikslų modelyje (žr. 1.7 pav. *Tikslų modelis*) galime pastebėti, kad informacijos paieškos ontologijų saugykloje gerinimas būtų pasiektas sugrupavus panašius individus ir automatiškai atpažinus panašius ontologijos individus.



1.7 pav. Tikslų modelis

Žemiau pateiktame modelyje (žr. 1.8 pav. *Veiklos konceptai*) matome pavaizduotus veiklos konceptus.



1.8 pav. Veiklos konceptai

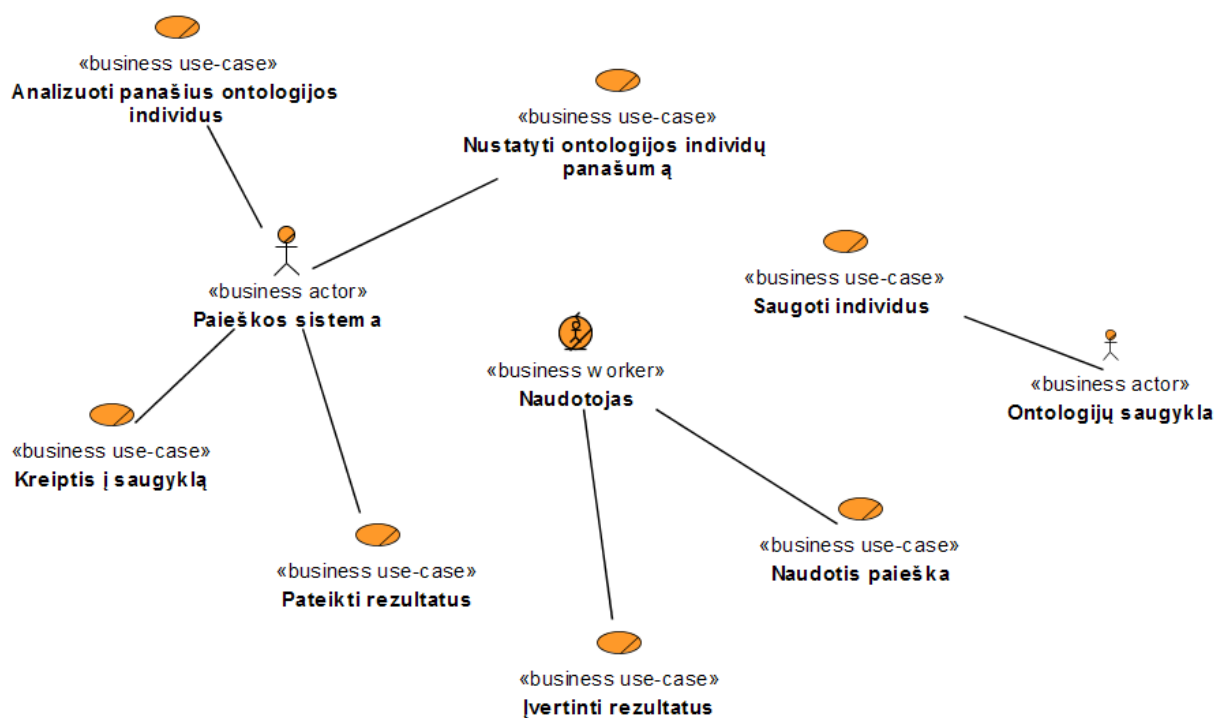
Veiklos kontekstas pateiktas 1.9 pav.



1.9 pav. Veiklos kontekstas

1.10 pav. pateikta veiklos panaudojimo atvejų diagrama.

Paieškos sistema gali analizuoti panašius ontologijos individus, nustatyti jų panašumą, pateikti rezultatus ir kreiptis į saugyklą. Naudotojas gali naudotis paieška ir įvertinti rezultatus. Ontologijų saugykla saugo individus.



1.10 pav. Veiklos panaudojimo atvejų diagrama

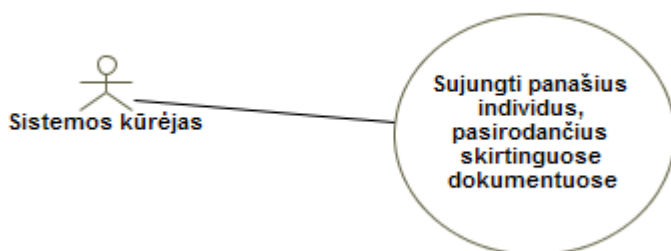
1.8. Analizės išvados

1. Išanalizuoti esami problemos sprendimo būdai, kitų autorių literatūra parodė, kad esami sprendimai turi trūkumą.
2. Tikslui pasiekti galėtų būti naudojamas Damerau–Levenshtein'o atstumas.
3. Atliktos analizės rezultatų pagrindu apibrėžtas siekiamas tyrimo problemos sprendimas - sukurti algoritmą, kuris sujungtų panašius ontologijos individus.

2. PANAŠIŲ ONTOLOGIJOS INDIVIDŲ ATPAŽINIMO IR SUJUNGIMO REIKALAVIMŲ SPECIFIKACIJA IR PROJEKTAS, FORMALUS APRAŠAS

2.1. Reikalavimų specifikacija

UML kompiuterizuojamų panaudojimo atvejų diagrama pateikta 2.1 paveiksle. Sistemos kūrėjas, pasinaudodamas programiniu įrankiu, gali sujungti panašius individus, pasirodančius skirtinguose dokumentuose.



2.1 pav. Kompiuterizuojamų PA diagrama

PANAUDOJIMO ATVEJAI

Panaudojimo atvejo „Susieti panašius individus, pasirodančius skirtinguose dokumentuose“ specifikacija pateikiama 2.1 lentelėje.

2.1 lentelė. Panaudojimo atvejo „Sulieti panašius individus, pasirodančius skirtinguose dokumentuose“ specifikacija

PA „Susieti panašius individus, pasirodančius skirtinguose dokumentuose“.	
Tikslas. Sulieti panašius individus.	
Aprašymas. Sistemos kūrėjas paleidžia panašių individų sujungimą.	
Prieš sąlyga	Sistema laukimo būsenoje
Aktorius	Sistemos kūrėjas
Sužadinimo sąlyga	Naudotojas nori susieti panašius individus
Susiję panaudojimo atvejai	Išplečia PA
	Apima PA
	Specializuoja PA
Pagrindinis įvykių srautas	Sistemos reakcija ir sprendimai
1. Spaudžiamas mygtukas „Pradėti sujungimą“	Paleidžiamas panašių individų suliejimas
Po sąlyga	Pasirodo sujungtų panašių individų sąrašas

NEFUNKCINIAI REIKALAVIMAI

2.2 lentelė. Nefunkcinis reikalavimas 1

Reikalavimo Nr.:	1	Reikalavimo tipas:	-
Aprašymas:	Sistemoje turi būti naudojamas sistemos šriftas		
Pagrindimas:	Visi tekstai sistemoje turi būti vienodai lengvai skaitomi.		

Šaltinis:	Užsakovas		
Tikimo kriterijus:	Visi tekstai rodomi naudotojo nustatytu šriftu.		
Naudotojo patenkinimas:	1	Naudotojo nepatenkinimas:	3
Priklausomybės:	-	Konfliktai:	-
Papildoma medžiaga:	-		
Istorija:	Sukurta 2014-05-10		

2.3 lentelė. Nefunkcinis reikalavimas 2

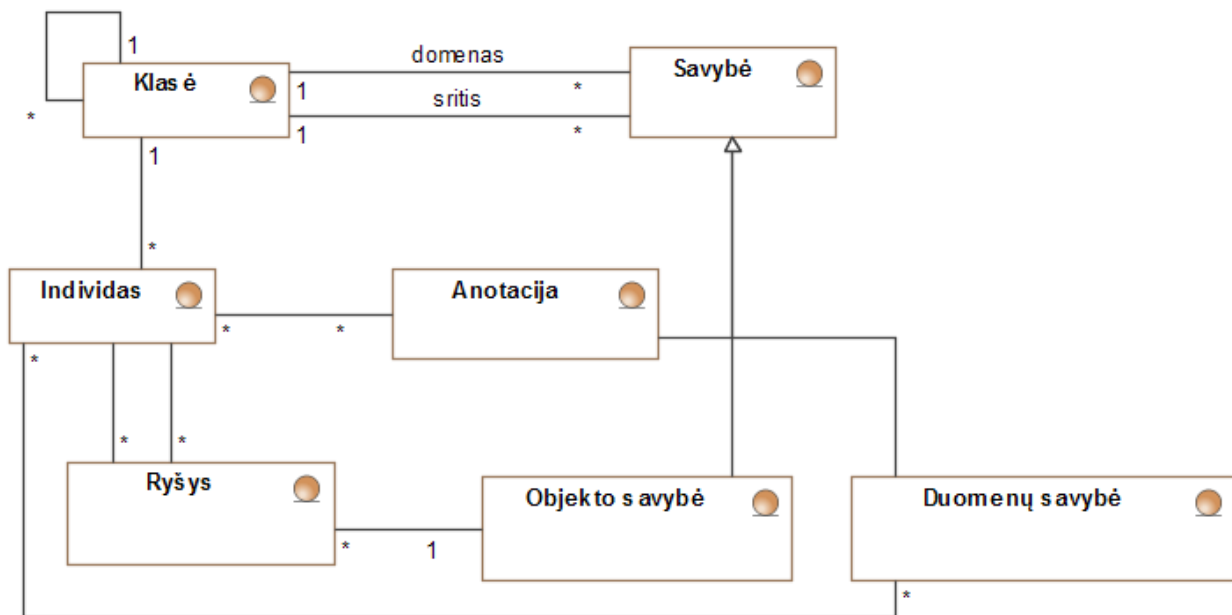
Reikalavimo Nr.:	2	Reikalavimo tipas:	-
Aprašymas:	Sistemoje turi būti naudojami standartiniai valdymo elementai		
Pagrindimas:	Naudotojai pažįstami su standartiniais valdymo elementais, todėl bus lengviau apmokyti juos dirbti su sistema.		
Šaltinis:	Užsakovas		
Tikimo kriterijus:	Nestandardiniai valdymo elementai naudojami tik ten, kur nėra galimybės juos pakeisti standartiniais.		
Naudotojo patenkinimas:	3	Naudotojo nepatenkinimas:	3
Priklausomybės:	-	Konfliktai:	-
Papildoma medžiaga:	-		
Istorija:	Sukurta 2014-06-14		

2.4 lentelė. Nefunkcinis reikalavimas 3

Reikalavimo Nr.:	3	Reikalavimo tipas:	-
Aprašymas:	Sistema turi būti pasiekiamą 99% laiko		
Pagrindimas:	Sistemos pasiekiamumas svarbus tam, kad būtų galima dirbti bet kada prireikus.		
Šaltinis:	Užsakovas		
Tikimo kriterijus:	Maksimalus sistemos nepasiekiamumo laikas – 4 dienos per metus.		
Naudotojo patenkinimas:	3	Naudotojo nepatenkinimas:	5
Priklausomybės:	-	Konfliktai:	-
Papildoma medžiaga:	-		
Istorija:	Sukurta 2014-06-14		

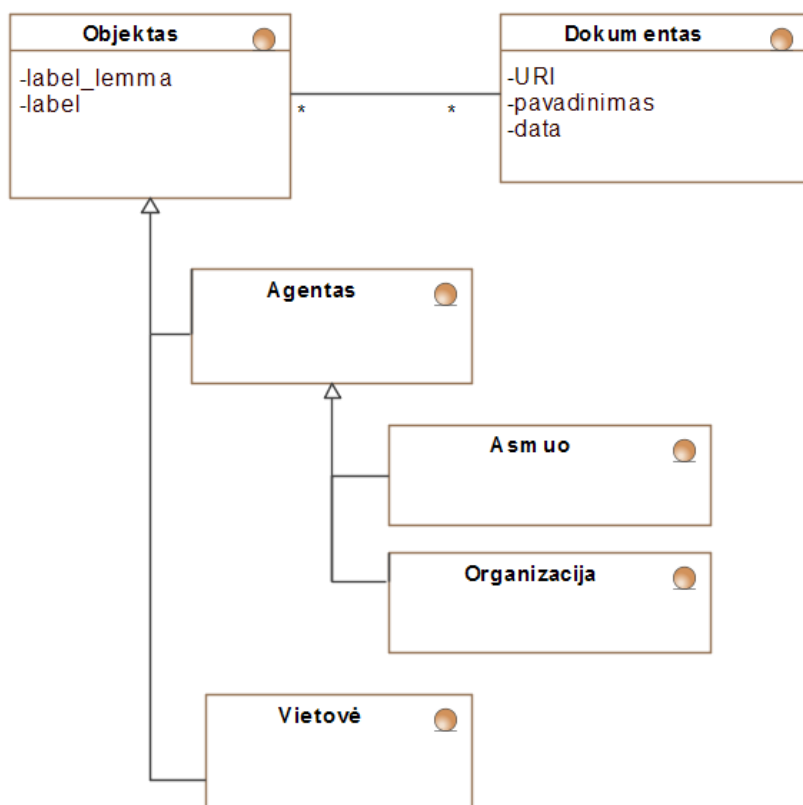
2.2. Dalykinės srities modelis

Nagrinėjamos ontologijos meta modelis pavaizduotas 2.2 pav.



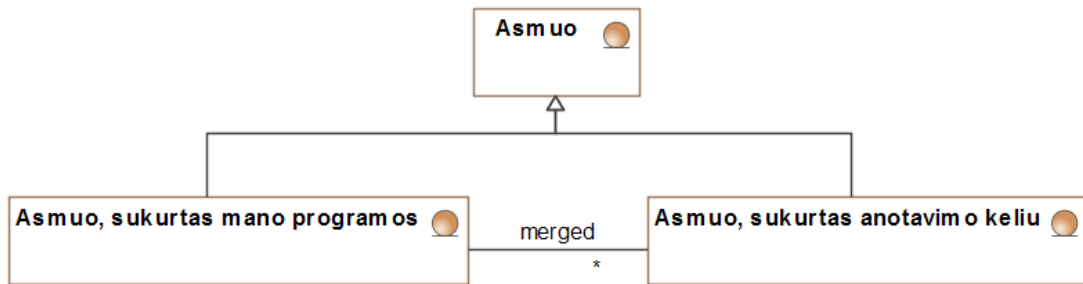
2.2 pav. Ontologijos meta modelis

Žemiau pateikta esybių klasių diagrama detalizuoja dalykinės srities modelį (žr. 2.3 pav. *Dalykinės srities modelis*).



2.3 pav. Dalykinės srities modelis

Žemiau pateikiamas asmenų klasę detalizuojantis modelis, kuriame *Asmuo, sukurtas mano programos* – tai asmuo, kurį sukūrė mano programa, kad apjungtų panašius asmenų individus, o *Asmuo, sukurtas anotavimo keliu* – tai teksto analizatorių atpažinti tekstuose paminėti asmenys.



2.4 pav. Asmenų klasę detalizuojantis modelis

2.3. Reikalavimų apibendrinimas

1. Šiame skyriuje buvo aptarti funkciniai ir nefunkciniai reikalavimai programinio įrankio kūrimui. Pagrindinė įrankio funkcija yra susieti panašius individus, pasirodančius skirtinguose dokumentuose.
2. Funkcinis reikalavimas buvo pateikti kompiuterizuojamų panaudojimo atvejų diagrama (2.1 pav.) ir specifikavimo lentelė (2.1 lentelė).
3. Nefunkciniai reikalavimai pateikti (žr. 2.2 – 2.4 lenteles) kuriamam įrankiui.

3. PANAŠIŲ ONTOLOGIJOS INDIVIDŲ ATPAŽINIMO IR SUJUNGIMO ALGORITMAS

Panašių ontologijos individų atpažinimo ir sujungimo algoritmui realizuoti buvo nuspręsta sukurti dvi taisykles, pagal kurias sujungiami rasti panašūs ontologijos individai:

1. Suliejimas pagal vardą ir pavardę, kai pavardė pilna, o vardas sutrumpintas:

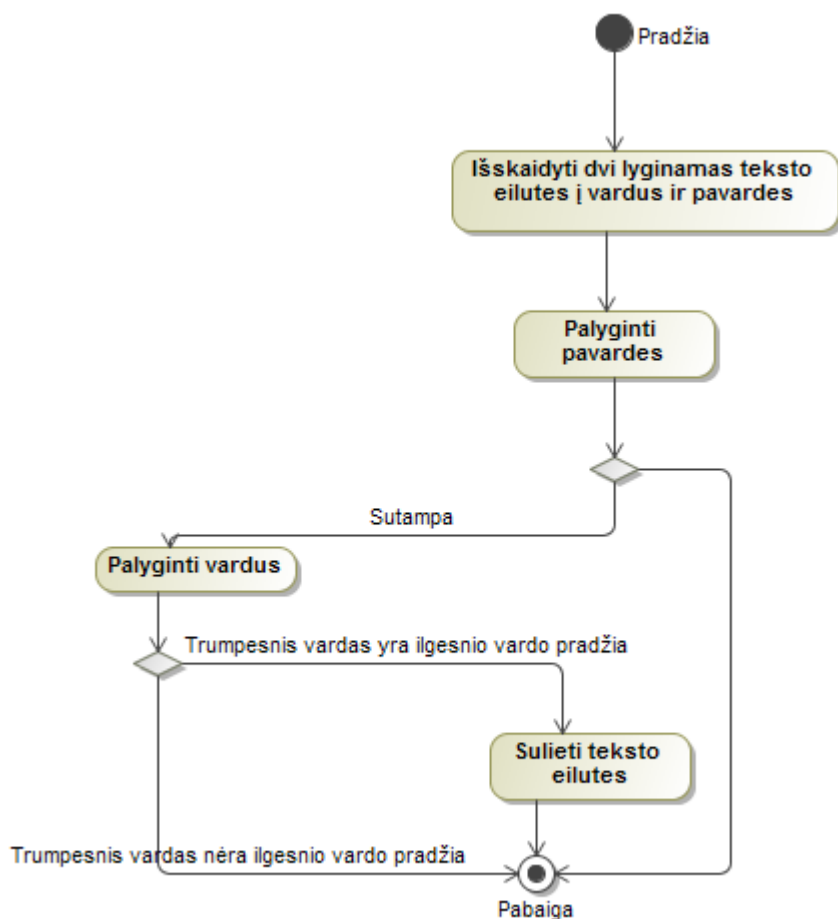
```
package com.ernesta.rules;
import org.apache.commons.lang3.StringUtils;
import com.ernesta.model.PersonRule;

public class FirstNameContractionRule implements PersonRule {

    @Override
    public boolean accept(String first, String second) {
        String firstFirstName = getFirstName(first);
        String firstLastName = getLastName(first);
        String secondFirstName = getFirstName(second);
        String secondLastName = getLastName(second);
        if (StringUtils.isNoneBlank(firstFirstName, firstLastName, secondFirstName,
secondLastName)) {
            if (firstLastName.equals(secondLastName)) {
                if (firstFirstName.endsWith(".") || secondFirstName.endsWith("."))
                {
                    String shorterFirstName = (firstFirstName.length() < secondFirstName.length())
? firstFirstName : secondFirstName;
                    String longerFirstName = (firstFirstName.length() < secondFirstName.length()) ?
secondFirstName : firstFirstName;
                    if (longerFirstName.endsWith(".")) {
                        return false;
                    }
                    if (longerFirstName.contains(shorterFirstName.replace(".", ""))) {
                        return true;
                    }
                }
            }
        }
        return false;
    }

    private String getFirstName(String name) {
        String parts[] = StringUtils.split(name.toLowerCase());
        if (parts.length <= 1) {
            return null;
        }
        return parts[0];
    }

    private String getLastName(String name) {
        String parts[] = StringUtils.split(name.toLowerCase());
        if (parts.length <= 1) {
            return null;
        }
        return parts[parts.length - 1];
    }
}
```



3.1 pav. Suliejimo pagal vardą ir pavardę, kai pavardė pilna, o vardas sutrumpintas algoritmas

Šis algoritmas (žr. 3.1 pav.) išskaido dvi lyginamas teksto eilutes į vardus ir pavardes. Išskaidymo rezultate gaunami du vardai ir dvi pavardės. Tada algoritmas lygina pavardes, žiūri ar jos sutampa, jeigu sutampa – žiūri, ar trumpesnis vardas yra ilgesnio pradžia. Jeigu trumpesnis vardas yra ilgesnio pradžia, tai lyginamos teksto eilutės suliejamos ir laikomos panašiomis.

2. Suliejimas, kai sutampa individų vardai ir pavardės:

```

package com.ernesta.rules;
import org.apache.commons.lang3.StringUtils;
import com.ernesta.model.PersonRule;

public class NameEqualityRule implements PersonRule {

    @Override
    public boolean accept(String first, String second) {
        String firstFirstName = getFirstName(first);
        String firstLastName = getLastName(first);
        String secondFirstName = getFirstName(second);
        String secondLastName = getLastName(second);
        if (StringUtils.isNoneBlank(firstFirstName, firstLastName, secondFirstName,
secondLastName)) {
            if (firstFirstName.equals(secondFirstName) &&
firstLastName.equals(secondLastName)) {
                return true;
            }
        }
        return false;
    }

    private String getFirstName(String name) {

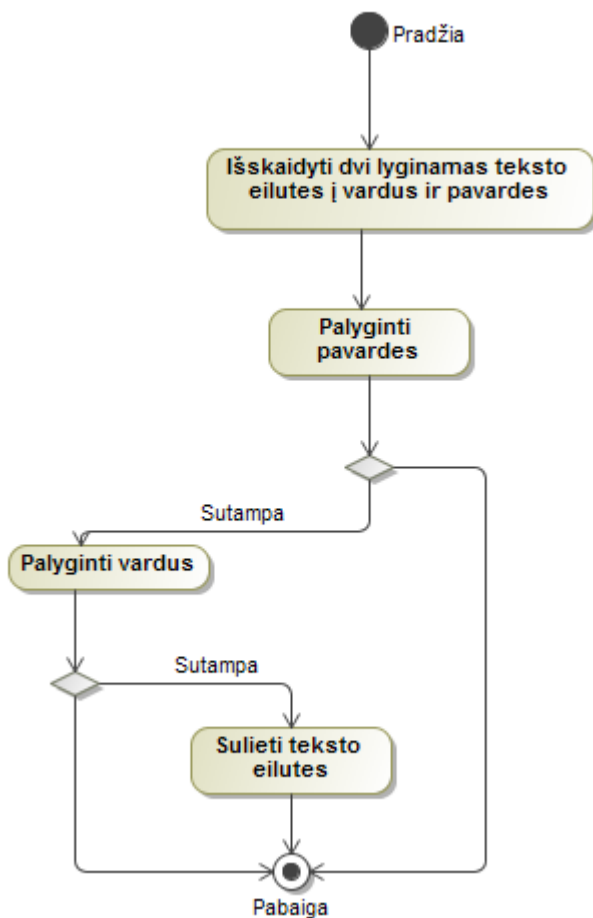
```

```

String parts[] = StringUtils.split(name.toLowerCase());
if (parts.length <= 1) {
    return null;
}
return parts[0];
}

private String getLastName(String name) {
String parts[] = StringUtils.split(name.toLowerCase());
if (parts.length <= 1) {
    return null;
}
return parts[parts.length - 1];
}
}

```



3.2 pav. Suliejimo, kai sutampa individų vardai ir pavardės, algoritmas

Algoritmas (žr. 3.2 pav.) išskaido dvi lyginamas teksto eilutes į vardus ir pavardes. Išskaidymo rezultate gaunami du vardai ir dvi pavardės. Tada algoritmas lygina pavardes, žiūri ar jos sutampa, jeigu sutampa – žiūri, ar sutampa vardai. Jeigu sutampa, tai lyginamos teksto eilutės suliejamos ir laikomos panašiomis.

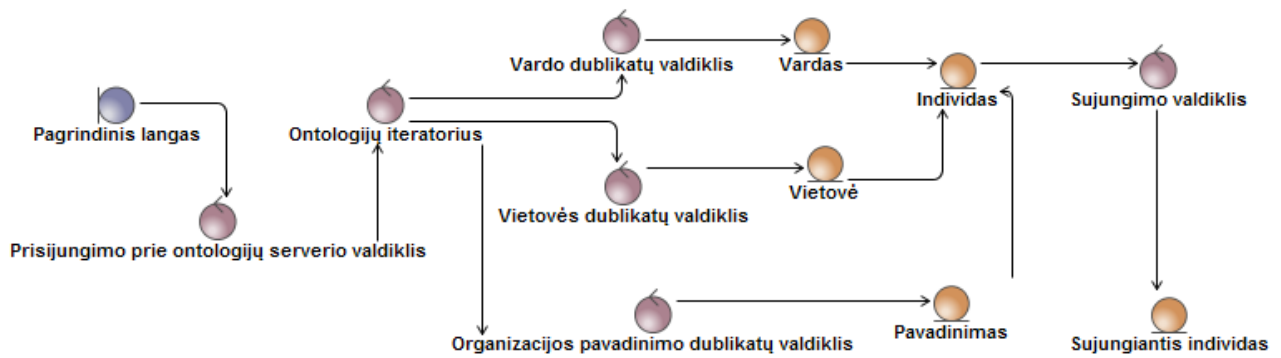
Sprendimui įgyvendinti panaudota OWL API biblioteka, palengvinanti ontologijų įrašymą.

3.1. Sistemos architektūra

Šiame poskyryje pateikiami reikalavimų analizės, loginė visos sistemos architektūros, naudotojo sąsajos klasių, veiklos logikos klasių modeliai.

3.1.1. Reikalavimų analizė

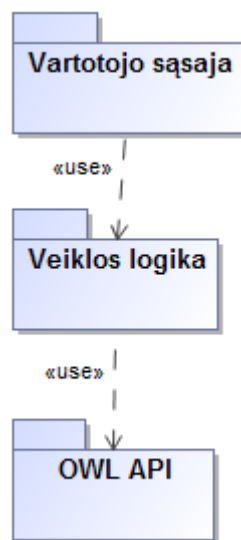
3.3 pav. pateikta analizės (robastiškumo) diagrama.



3.3 pav. Analizės (robastiškumo) diagrama

3.1.2. Loginė visos sistemos architektūra

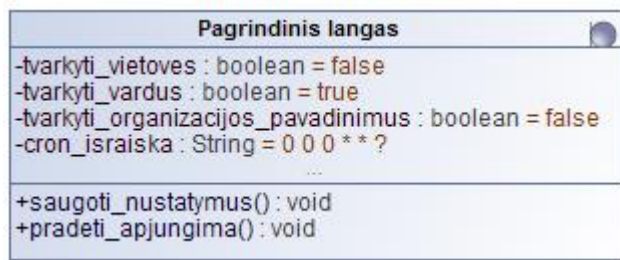
Loginė visos sistemos architektūra pateikta 3.4 pav.



3.4 pav. Loginė visos sistemos architektūra

3.1.3. Naudotojo sąsajos klasių modelis

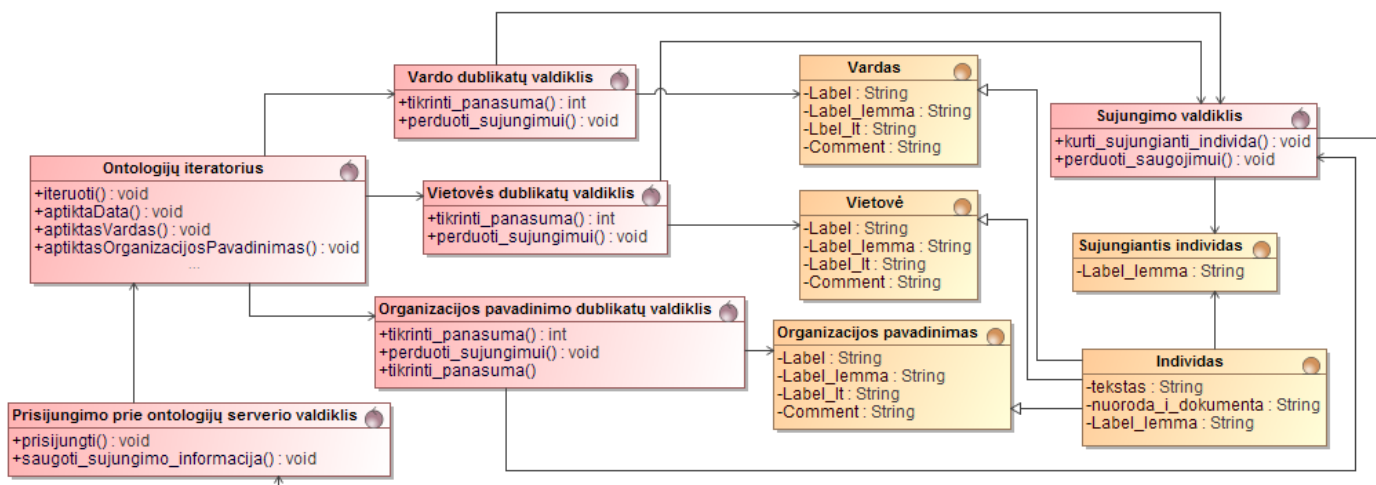
Naudotojo sąsajos klasių modelis nurodytas 3.5 pav.



3.5 pav. Naudotojo sąsajos klasių modelis

3.1.4. Veiklos logikos klasių modelis

Veiklos logikos klasių modelis pateiktas 3.6 pav.

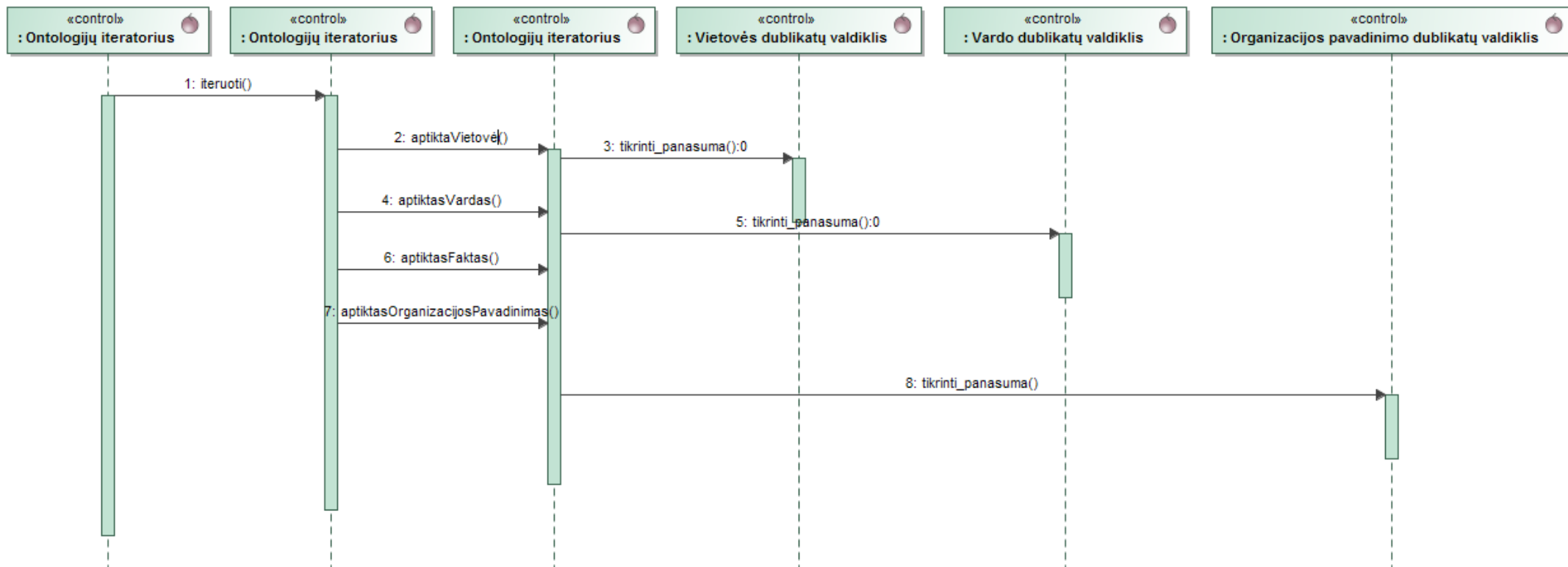


3.6 pav. Veiklos logikos klasių modelis

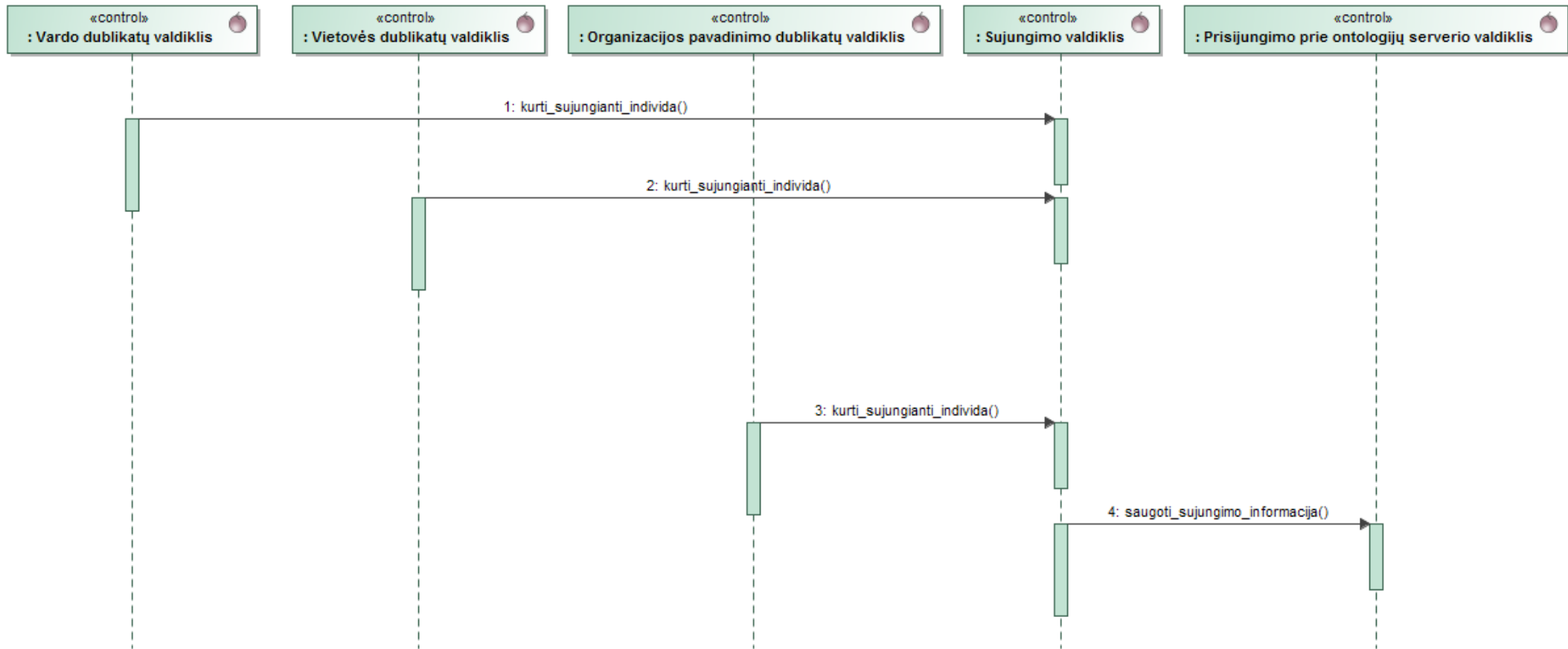
3.2. Sistemos elgsenos modelis

Panaudojimo atvejų diagramos, kurios parodo bendravimą tarp klasių, valdiklių ir langų yra pateikiamos šiame skyriuje.

Naudotojo sąsajos lange, paspaudus mygtuką „Pradėti sujungimą“, paleidžiamas panašių individų suliejimas, kuris apima tiek panašių individų suradimą, tiek jų sujungimą. Suliejimas išskaidytas į dvi sekas. Paspaudus mygtuką vykdomas panašių individų suradimas (žr. 3.7 pav. *Panašių individų suradimas*), po kurio vykdomas jų suliejimas (žr. 3.8 pav. *Panašių individų suliejimas*).

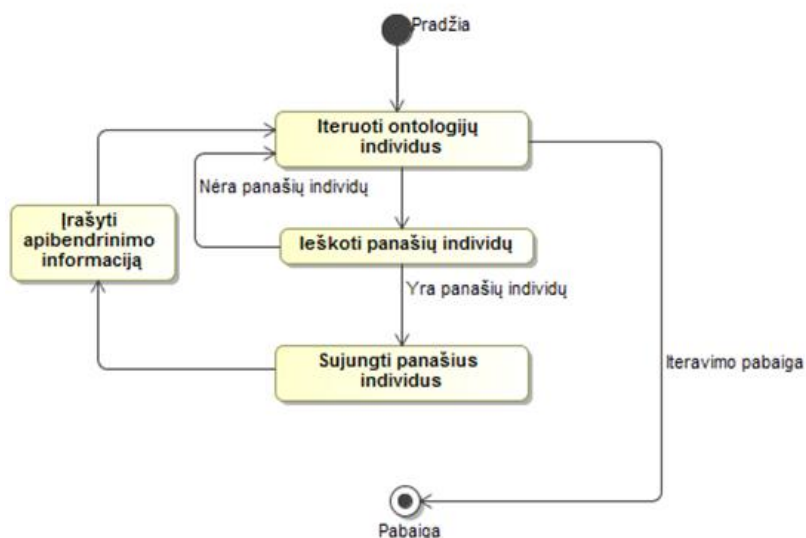


3.7 pav. Panašių individų suradimas



3.8 pav. Panašių individų suliejimas

Veiklos diagrama, naudojant būsenų mašinos elementus, pateikta 3.9 pav.



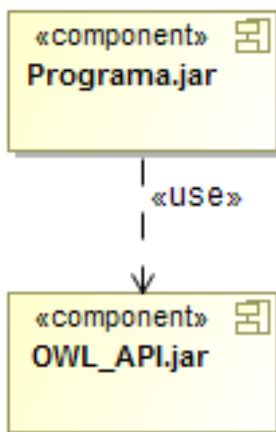
3.9 pav. Veiklos diagrama, naudojant būsenų mašinos elementus

3.3. Realizacijos modelis

Šiame poskyryje pateikiami programinių komponentų architektūros ir diegimo modeliai.

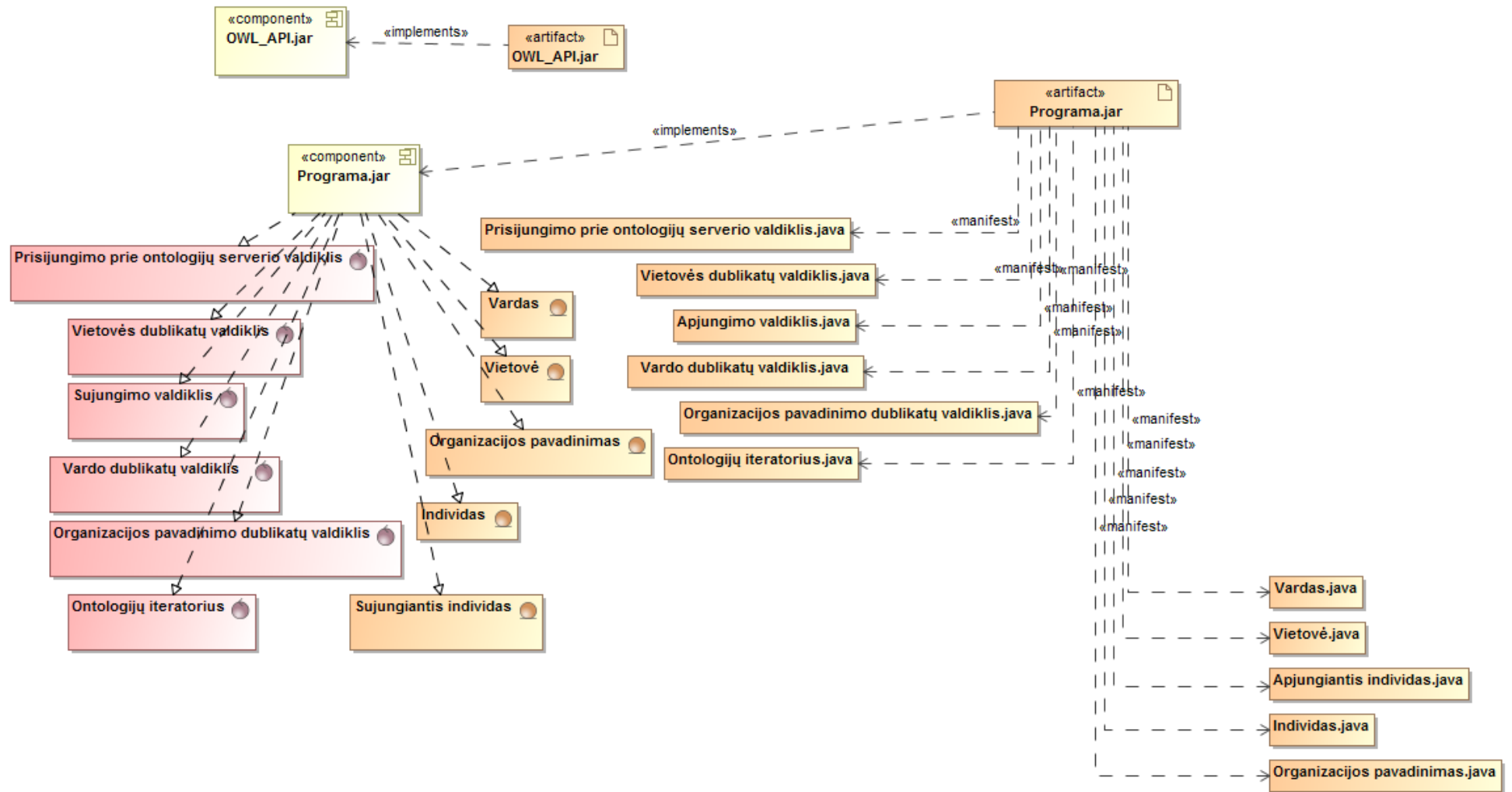
3.3.1. Programinių komponentų architektūra

Programinių komponentų architektūra pateikta 3.10 pav.



3.10 pav. Programinių komponentų architektūra

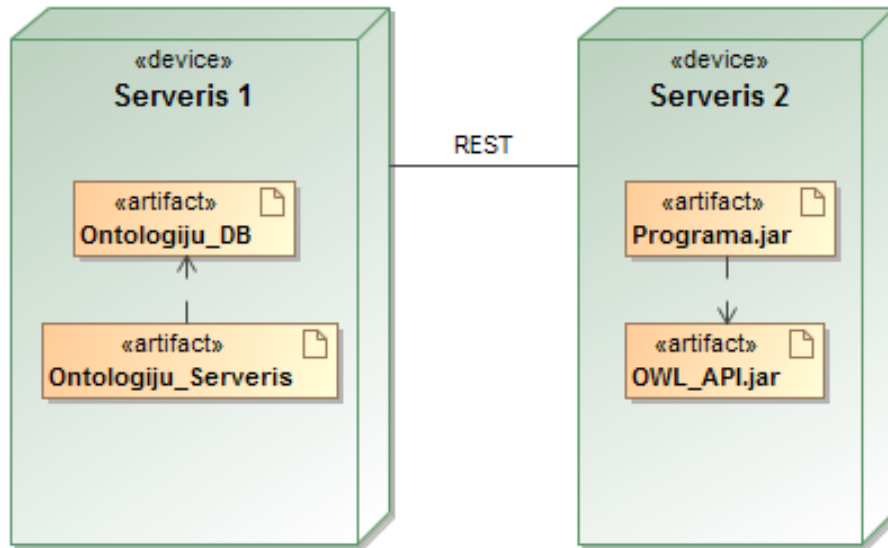
Komponentų realizacija artefaktais pateikta žemiau esančioje schemeje (žr. 3.11 pav.).



3.11 pav. Komponentų realizacija artefaktais

3.3.2. Diegimo modelis

Diegimo modelis pateiktas 3.12 pav.



3.12 pav. Diegimo modelis

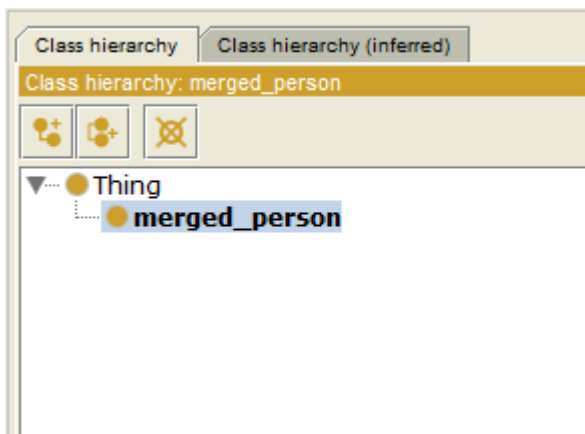
4. SPRENDIMO REALIZACIJA IR TESTAVIMAS

4.1. Sprendimo realizacijos ir veikimo aprašas

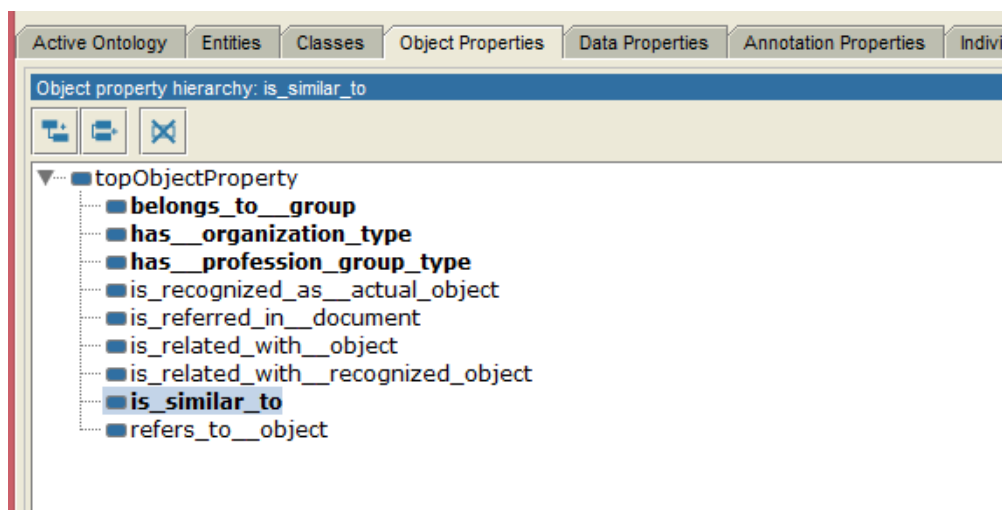
Realizavimo technologijos:

- Programa prašyta naudojant Java kalbą ir Java 8 standartą.
- Grafinė aplinkai buvo naudojama JavaFX, nes yra naujausia grafinės vartotojo sąsajos technologija Java kalbai.
- Dialogams buvo naudojama ControlsFX biblioteka.
- Darbui su simbolių eilutėmis buvo naudojama Apache Commons Lang biblioteka, nes ji palengvina darbą su simbolių eilutėmis.
- Darbui su ontologijomis buvo naudojama OWL API biblioteka, nes ji turi visą reikalingą funkcionalumą uždaviniui spręsti ir aiškią dokumentaciją.
- Programos testavimui buvo naudojamas JUnit testavimo karkasas, nes tai vienas plačiausiai naudojamų vienetų testavimo karkasų Java kalbai.
- Programos surinkimui (kompiliavimui) ir priklausomybių valdymui buvo naudojamas Maven karkasas, nes jis gerokai palengvina visą šį procesą.

Sprendimo realizacijos metu buvo sukurta nauja schema *MergedObjects* (žr. 4.1 pav.), kurioje yra esybė *merged_person* ir papildyta *Agents* schema (žr. 4.2 pav.), į kurią įdėta savybė *is_similar_to*.



4.1 pav. MergedObjects schema

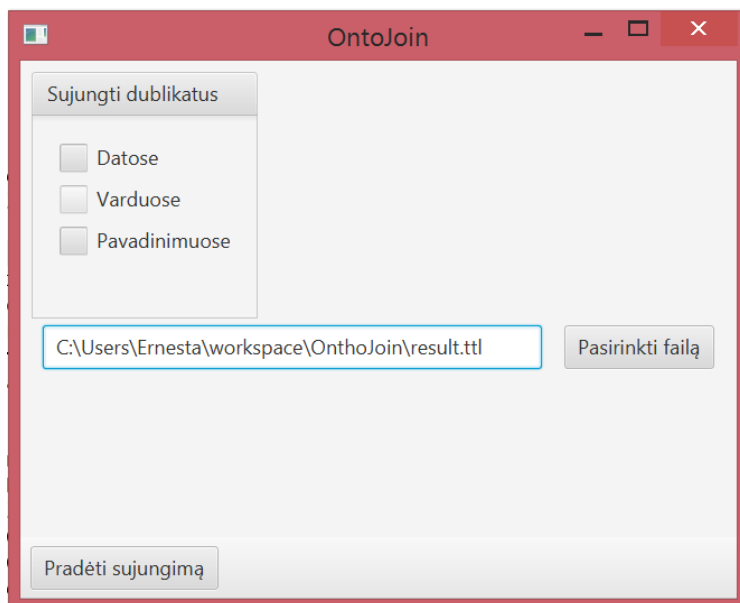


4.2 pav. Papildyta Agents schema

Sukurtos dvi ontologijų sujungimo taisyklės: sujungimas pagal vardą ir pavardę, kai pavardė pilna, o vardas sutrumpintas ir sujungimas, kai sutampa individų vardai ir pavardės (žr. [3 skyrių: Panašių ontologijų individų atpažinimo ir sujungimo algoritmas](#)).

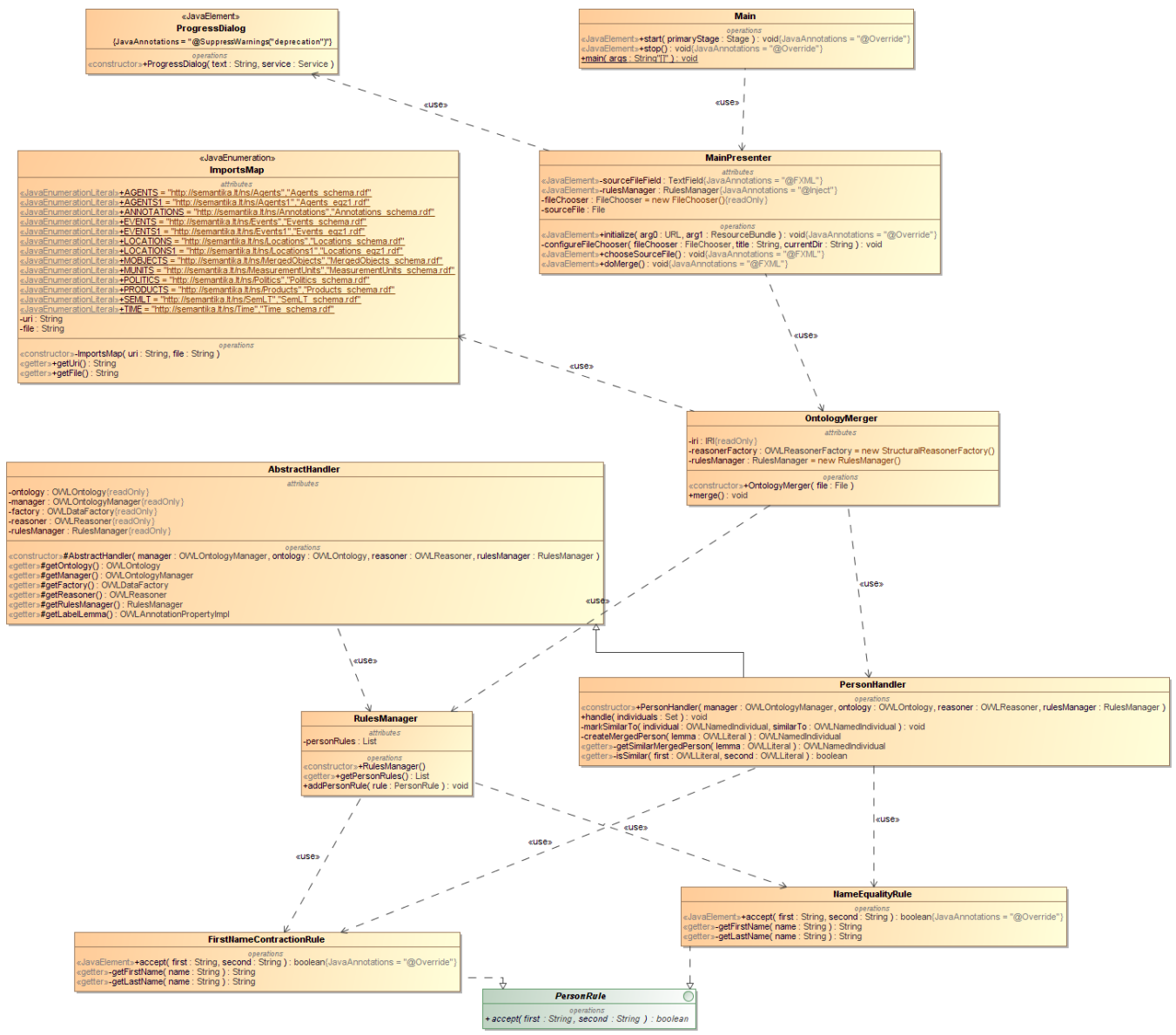
Sistemos veikimas:

Pradiniame programos lange (žr. 4.3 pav. *Vartotojo sąsaja pagrindiniame programos lange*) sistemos naudotojai pasirenka ontologijos failą ir paspaudžiamas mygtukas „Pradėti sujungimą“. Failas yra nuskaitymas. Išskviečiamas taisyklių vadovas, pridedamos taisyklės, užkraunamos ontologijos iš dokumento. Pagal label_lemma's nustatomi individai, kurie sujungiami pagal panašumą. Jie yra lyginami taikant dvi taisykles: susiejimą pagal vardą ir pavardę, kai pavardė pilna, o vardas sutrumpintas ir susiejimą, kai sutampa individų vardai ir pavardės. Rasti panašūs ontologijos individai susiejami. Gauta ontologija įrašoma į naują failą. Rezultatai peržiūrimi su *Protege* programa.



4.3 pav. Vartotojo sąsaja pagrindiniame programos lange

Įrankio klasių modelis pateiktas 4.4 pav., jis apibrėžia statinę sistemos struktūrą. Pagrindiniai diagramos elementai yra klasės ir asociacijos tarp jų. Klasės apibūdina objektų aibę, o asociacija apibūdina ryšių (angl. *link*) aibę.



4.4 pav. Įrankio klasių modelis

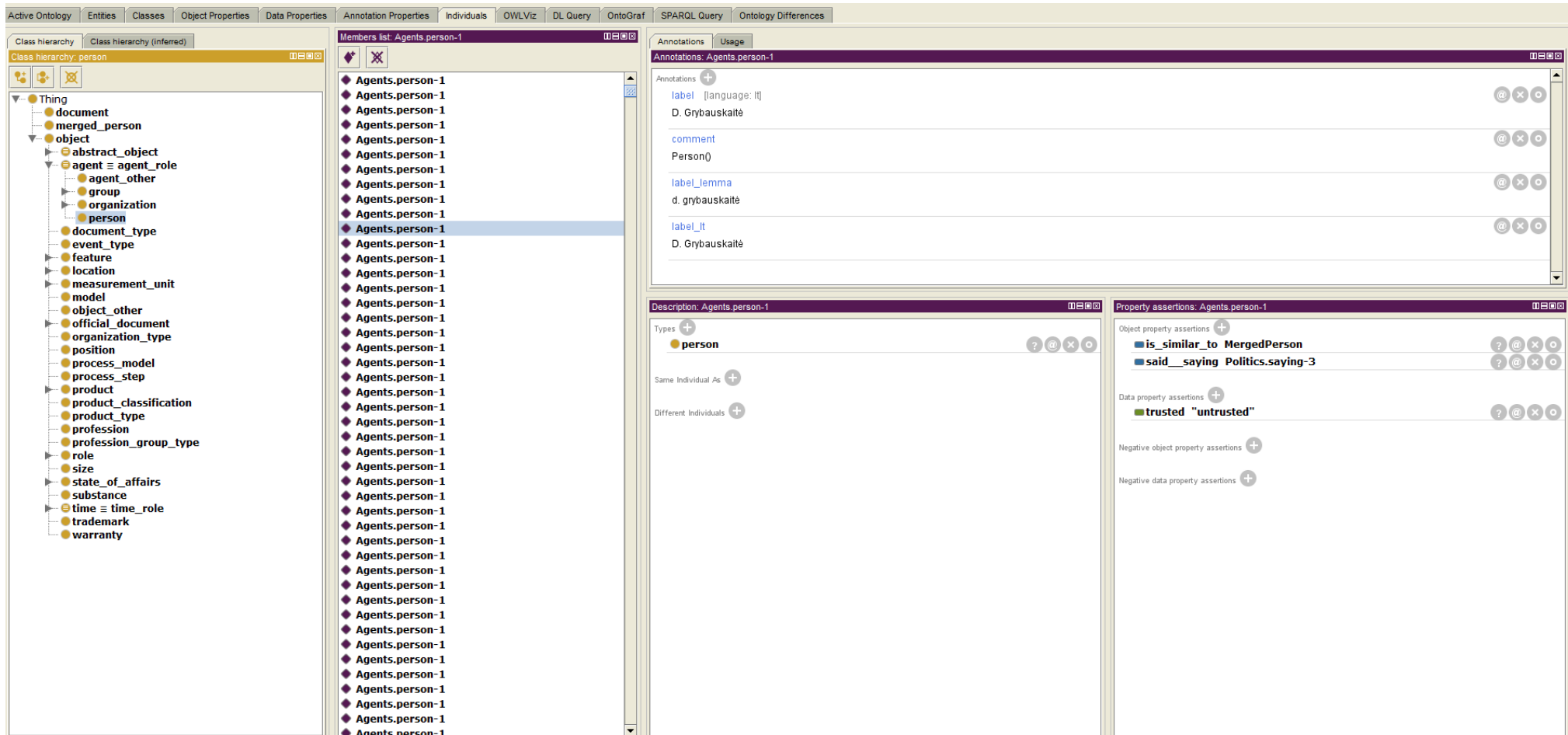
Ontologijoje ieškant asmens, pavyzdžiui, Gediminas Kirkilas, bus pereinama per visus tikruosius ontologijoje esančius asmenų įrašus, tada pagal ieškomą žodį bus surandamas asmuo, atitinkantis paiešką. Paieškos sistema ras susietą apibendrinantį asmenį ir tada galės atvaizduoti visus įrašus, kuriuos jis apibendrina.

Sukurta sprendimas naudingas, nes problema, kilusi dėl besidubliuojančių ontologijos individų – išspręsta. Algoritmas automatiškai atpažįsta panašius ontologijos individus, sumažina ontologijos individų pasikartojimą, susieja juos pagal panašumą.

4.2. Testavimo modelis, duomenys, rezultatai

Sprendimas testuojamas su ontologija, kurioje visi pavadinimai yra lietuvių kalba.

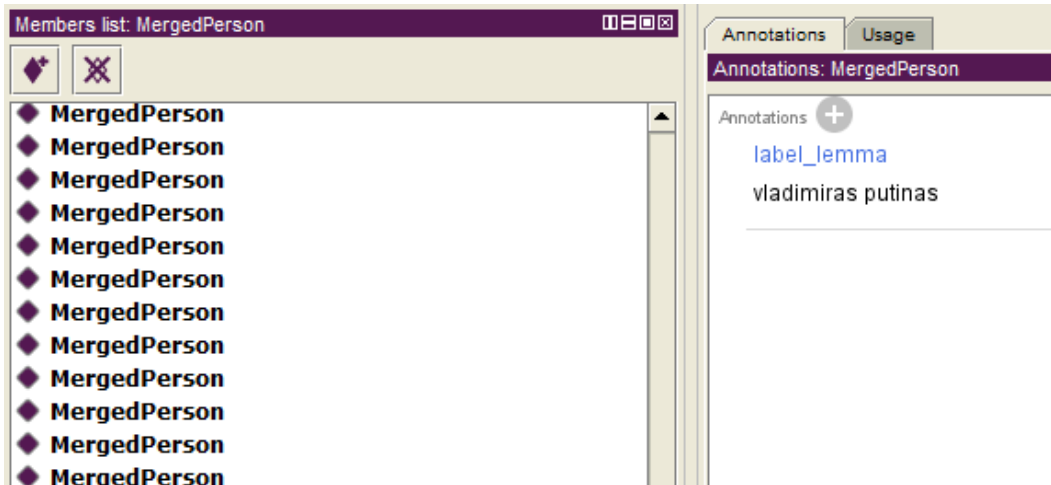
Testiniai duomenys faile: semantics_dump_unmerged.ttl (žr. 4.5 pav. Vartotojo sąsaja pagrindiniame programos lange).



4.5 pav. Vartotojo sąsaja pagrindiniame *Protege* programos lange

Rezultatai:

Pasirinkta, nuskaityta ontologija įrašoma į rezultatų failą. Rezultatai peržiūrimi su *Protege* programa. Pagal `Label_lemma`'s nustatyti individai sujungiami pagal panašumą, taip sumažinant ontologijos individų pasikartojimą. Pavyzdžiui, asmuo Vladimiras Putinas (žr. 4.6 pav. *Bendrinė žodžio forma, nusakanti individą*) ontologijoje yra paminėtas kelis kartus. 4.7 pav. pateikiami atpažinti panašūs ontologijos individai į Vladimirą Putiną.



4.6 pav. Individą nuskanti bendrinė žodžio forma

The screenshot displays an ontology editor interface. On the left, a list of 25 instances of the class 'MergedPerson' is shown, with the 15th instance highlighted. The right pane is divided into three sections: 'Annotations' (empty), 'Usage' (showing 16 uses of 'MergedPerson' with 'is_similar_to' assertions), and 'Description' (showing the type 'merged_person' and options for 'Same Individual As' and 'Different Individuals'). The bottom right pane shows 'Property assertions' for 'MergedPerson', including 'Object property assertions', 'Data property assertions', 'Negative object property assertions', and 'Negative data property assertions'.

4.7 pav. Ontologijoje rasti ir apjungti panašūs individai

5. EKSPERIMENTINIS SPRENDIMO (PANAŠIŲ ONTOLOGIJOS INDIVIDŲ ATPAŽINIMO IR SUJUNGIMO) TYRIMAS

5.1. Eksperimento planas

Eksperimento metu bus parašytos SPARQL užklauskos, kurias įvykdžius bus išvedami panašūs asmenys su pilnais vardais ir nuorodomis į dokumentą. Užklauskos vykdomos su *Protege* programa.

Pagal SPARQL užklauską bus pereinama per visus tikruosius ontologijoje esančius asmenų įrašus, tada pagal ieškomą žodį bus surandamas asmuo, atitinkantis paiešką. Paieškos sistema ras susietą apibendrinantį asmenį pagal savybę *similar to* ir tada galės atvaizduoti visus įrašus, kuriuos jis apibendrina pagal savybę *similar to*.

Eksperimento kontekstas – tai tekstai, kuriuose užrašyti įvairūs žmonių vardai ir pavardes, pvz., žiniasklaidos straipsniai ir pan.

Eksperimento metu siekiama įvertinti algoritmo kokybę ir išsiaiškinti, ar sudaryta galimybė automatiškai nustatyti panašius ontologijos individus, juos susieti ir tokiu būdu sumažinti informacijos ontologijoje dubliavimą.

5.2. Eksperimento rezultatai

Iš analizuojamos ontologijos buvo išrinkti dokumentai, kurie susieti su tam tikrais asmenimis. Pagal šaltinio nuorodą internete surasti analizuoti straipsniai. Atsidarius straipsnį matoma, kokie individai buvo atpažinti.

SPARQL užklausa įrašoma *Protege* programos kortelėje „SPARQL Query“. Žemiau pateikta SPARQL užklausa išrenka iš automatiškai išanalizuotų internete publikuotų lietuviškų straipsnių tekstų individų pavadinimus, kurie panašūs į ieškomą individą „dalia“. Rezultatuose (žr. 5.1 pav.) išvedamos visais linksniais ir sutrumpintais vardais Dalios, kurios buvo paminėtos išanalizuotuose straipsniuose. Šalia paminėto asmens nurodomas to straipsnio URL, kuriame buvo paminėta Dalia.

Įvykdžius užklauską parodomi visi (pasikartojantys) įrašai.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?uri
WHERE {
  ?subject a ag:person .
  ?subject an:label_lemma ?label_lemma .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "dalia", "i")) .
  ?subject ag:is_similar_to ?merged .
  ?found ag:is_similar_to ?merged .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
}
```

found_label	uri
"Dalia Grybauskaitė"@lt	"http://www.balsas.lt/naujiena/760570/es-tikslas-prieinama-saugi-ir-bendra-energetikos-rinka"@
"D. Grybauskaitės"@lt	"http://www.balsas.lt/naujiena/743184/politologai-nesutaria-del-d-grybauskaites-galimybiu-tapti-es-prezidente/2"@
"Dalia Grybauskaitė"@lt	"http://www.balsas.lt/naujiena/526472/del-d-kreivio-premjeras-su-partija-ir-koalicija-tarsis-kita-savaite"@
"Dalia Grybauskaitė"@lt	"http://www.valstietis.lt/Pradzia/Naujienos/Pasaulyje/ES-vaira-perimancios-Graikijos-premjeras-Kasandros-pranasyste-neissipilde"@
"D. Grybauskaitė"@lt	"http://www.delfi.lt/news/daily/lithuania/m-laurinavicius-v-janukovycius-butu-persigalvojes-vienu-atveju.d?id=63408928"@
"D. Grybauskaitė"@lt	"http://www.alfa.lt/straipsnis/15720176/grauziniene-valstybes-viduje-neturime-ieskoti-priesu-vardan-asmeniniu-interesu"@
"D. Grybauskaitė"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/30635/m_laurinavicius_ukrainos_prezidentas_butu_persigalvojes_tik_vienu_atveju"@
"D. Grybauskaitė"@lt	"http://www.balsas.lt/naujiena/764738/d-grybauskaite-ukrainos-valdzia-sali-veda-i-niekur?utm_source=susijeStr&utm_medium=susijeDienos-kronika&utm_campaign=naujiena"@
"D. Grybauskaitė"@lt	"http://www.delfi.lt/news/daily/lithuania/bkomorowski-vaziuoju-i-lietuva-palaikyti-gera-abieju-valstybiu-vadovu-lankymosi-tradicija.d?id=60688787"@
"Dalius Grybauskaitės"@lt	"http://www.delfi.lt/verslas/verslas/prezidentas-patareja-atskiroms-salims-sutarta-es-parama-nebus-derybu-su-ep-objektas.d?id=60896671"@
"Dalia Grybauskaitė"@lt	"http://www.valstietis.lt/Pradzia/Naujienos/Lietuvoje/ES-vadovai-neitikino-Ukrainos-del-asociacijos-sutarties"@
"D. Grybauskaitė"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai"@
"Dalia Grybauskaitė"@lt	"http://www.balsas.lt/naujiena/764442/politologas-apie-l-grauzinienes-keleone-i-ukraina-tai-kamikadzes-misija-politikei-ir-partijai"@
"Dalia Grybauskaitė"@lt	"http://www.alfa.lt/straipsnis/15840186/prezidentas-penkeriu-metu-ataskaita-pakvipo-rinkimu-kampanija"@
"Dalia Grybauskaitė"@lt	"http://www.balsas.lt/naujiena/629300/valdzia-istiko-krize-prezidente-nevertins-jokiu-koaliciono-sudarymo-galimybiu-kreipsis-i-kt-ir-lauks-jo-sprendimo"@
"Dalius Grybauskaitės"@lt	"http://www.diena.lt/naujienos/lietuva/politika/lietuvos-konservatoriai-remia-v-dombrovskio-kandidatura-vadovauti-europos-komisijai-615624"@
"D. Grybauskaitė"@lt	"http://www.valstietis.lt/Pradzia/Naujienos/Lietuvoje/ES-vadovai-neitikino-Ukrainos-del-asociacijos-sutarties"@
"Dalia Grybauskaitė"@lt	"http://www.valstietis.lt/Pradzia/Naujienos/Lietuvoje/ES-vadovai-neitikino-Ukrainos-del-asociacijos-sutarties"@
"Dalius Grybauskaitės"@lt	"http://www.kauno.diena.lt/naujienos/lietuva/politika/lietuvos-konservatoriai-remia-v-dombrovskio-kandidatura-vadovauti-europos-komisijai-615624"@
"D. Grybauskaitė"@lt	"http://www.delfi.lt/news/daily/lithuania/dgrybauskaite-per-anksti-spekuliuoti-del-jtymosenko.d?id=63146414"@
"Dalia Grybauskaitė"@lt	"http://www.delfi.lt/news/daily/lithuania/dgrybauskaite-per-anksti-spekuliuoti-del-jtymosenko.d?id=63146414"@

5.1 pav. SPARQL užklauso rezultatai ieškant individo „dalia“

Žemiau pateikta SPARQL užklausa išrenka iš automatiškai išanalizuotų internete publikuotų lietuviškų straipsnių tekstų individų pavadinimus, kurie panašūs į ieškomą individą „loreta“. Rezultatuose (žr. 5.2 pav.) išvedamos visais linksniais ir sutrumpintais vardais Loretos, kurios buvo paminėtos išanalizuotuose straipsniuose. Šalia paminėto asmens nurodomas to straipsnio URL, kuriame buvo paminėta Loreta.

DISTINCT raktažodis atrenka tik unikalius įrašus. Užklauso vykdyimo rezultatuose pasikartojantys įrašai nerodomi. Pavyzdžiui, jeigu Loreta Grauzinienė buvo tame pačiame straipsnyje ir ta pačia žodžio forma paminėta kelis kartus (nuorodos į dokumentą sutaps), tai ši užklausa rodyt tik vieną įrašą, kuriame buvo paminėta Loreta Grauzinienė su nuoroda į tą patį vieną dokumento URL.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT DISTINCT ?found_label ?uri
WHERE {
  ?subject a ag:person .
  ?subject an:label_lemma ?label_lemma .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "loreta", "i")) .
  ?subject ag:is_similar_to ?merged .
  ?found ag:is_similar_to ?merged .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
}
```

found_label	uri
"L. Grauzinienė"@lt	"http://www.balsas.lt/naujiena/771316/l-grauziniene-kaltina-lb-vadova-politikavimu?utm_source=kitiRubrikosStr&utm_medium=susijeVerslas&utm_campaign=naujiena"@
"L. Grauzinienė"@lt	"http://www.delfi.lt/news/daily/lithuania/premjer-as-pasigedo-l-grauzinienes-solidumo.d?id=63708010"@
"Loreta Grauzinienė"@lt	"http://www.balsas.lt/naujiena/771316/l-grauziniene-kaltina-lb-vadova-politikavimu"@
"Loreta Grauzinienė"@lt	"http://www.alfa.lt/straipsnis/191628/dp-neatskleidzia-kokiomis-lesomis-dengs-savo-rinkimu-kampanija"@
"L. Grauzinienė"@lt	"http://www.alfa.lt/straipsnis/191628/dp-neatskleidzia-kokiomis-lesomis-dengs-savo-rinkimu-kampanija"@
"L. Grauzinienė"@lt	"http://www.alfa.lt/straipsnis/15697653/valdanciuju-atsakas-prezidentei-lauksime-faktu-apie-kremliaus-itaka-darbieciams"@
"L. Grauzinienė"@lt	"http://www.balsas.lt/naujiena/764442/politologas-apie-l-grauzinienes-kelione-i-ukraina-tai-kamikadzes-misija-politikei-ir-partijai"@
"Loreta Grauzinienė"@lt	"http://www.alfa.lt/straipsnis/15720176/grauziniene-valstybes-viduje-neturime-ieskoti-priesu-wardan-asmeniniu-interesu"@
"L. Grauzinienė"@lt	"http://www.diena.lt/naujienos/lietuva/salies-pulsas/siaurusevicius-l-grauziniene-pasakojo-apie-finansus-ir-pasirengima-eurovizijai-424889"@
"L. Grauzinienė"@lt	"http://www.alfa.lt/straipsnis/15780192/grauziniene-kai-kurie-parlamentarai-zavejosi-putino-kalba"@
"L. Grauzinienė"@lt	"http://www.balsas.lt/naujiena/759512/cosac-dalyviams-parodytas-b-bradausko-isbrokuotas-filmas-apie-nepriklausomybes-akturima"@
"Loreta Grauzinienė"@lt	"http://www.delfi.lt/news/daily/lithuania/v-uspaskichas-apie-moteriska-l-grauzinienes-paslapti-nemaciau.d?id=62971200"@
"Loreta Grauzinienė"@lt	"http://www.delfi.lt/news/daily/lithuania/l-grauziniene-del-reveranso-rusams-pritariu-b-bradausko-sprendimui.d?id=63068794"@
"Loreta Grauzinienė"@lt	"http://kauno.diena.lt/naujienos/lietuva/politika/premjer-as-l-grauzinienes-kalbos-apie-aura-del-reitingu-609487"@
"Loretos Grauzinienės"@lt	"http://www.ve.lt/naujienos/lietuva/lietuvas-naujienos/ministru-atlyginimai-siemet-pakilo-apie-25-tukst-litu-1182260/"@
"L. Grauzinienė"@lt	"http://klaipeda.diena.lt/naujienos/lietuva/salies-pulsas/siaurusevicius-l-grauziniene-pasakojo-apie-finansus-ir-pasirengima-eurovizijai-424889"@
"Loretos Grauzinienės"@lt	"http://www.balsas.lt/naujiena/764442/politologas-apie-l-grauzinienes-kelione-i-ukraina-tai-kamikadzes-misija-politikei-ir-partijai"@
"Loreta Grauzinienė"@lt	"http://www.diena.lt/naujienos/lietuva/salies-pulsas/siaurusevicius-l-grauziniene-pasakojo-apie-finansus-ir-pasirengima-eurovizijai-424889"@
"Loreta Grauzinienė"@lt	"http://www.delfi.lt/news/daily/lithuania/premjer-as-pasigedo-l-grauzinienes-solidumo.d?id=63708010"@
"Loreta Grauzinienė"@lt	"http://www.balsas.lt/naujiena/771316/l-grauziniene-kaltina-lb-vadova-politikavimu?utm_source=kitiRubrikosStr&utm_medium=susijeVerslas&utm_campaign=naujiena"@
"L. Grauzinienė"@lt	"http://www.balsas.lt/naujiena/771316/l-grauziniene-kaltina-lb-vadova-politikavimu"@

5.2 pav. SPARQL užklauso rezultatai naudojant DISTINCT raktažodį

Sukurtas algoritmas ir parašytos SPARQL užklauso sudaro galimybę automatiškai nustatyti panašius ontologijos individus ir juos susieti. Tokiu būdu sumažinamas informacijos ontologijoje dubliavimas. Apie egzempliorius sukauptas išanalizuotas informacijos rezultatas yra atvaizduojamas ontologijoje, kad vykdant paiešką kaskart nereikėtų analizuoti ontologijos individų panašumo.

Žemiau pateikta SPARQL užklausa išrinko iš automatiškai išanalizuotų internete publikuotų lietuviškų straipsnių tekstų individų pavadinimus, kurie panašūs į ieškomą individą „obama“. Rezultatuose (žr. 5.3 pav.) išvedami visais linksniais, pilnais ir su sutrumpintais vardais Obamos, kurie buvo paminėti išanalizuotuose straipsniuose.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?uri
WHERE {
  ?subject a ag:person .
  ?subject an:label_lemma ?label_lemma .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i") ) .
  ?subject ag:is_similar_to ?merged .
  ?found ag:is_similar_to ?merged .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to _object ?found .
  ?document sem:document_URI ?uri .
}
```

Įvykdžius užklausą parodomi visi (pasikartojantys) įrašai.

found_label	uri
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva.d?id=62187757"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai"@
"B. Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai"@
"Barackas Obama"@lt	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva.d?id=62187757"@
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801"@
"B. Obama"@lt	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801"@
"B. Obama"@lt	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801"@
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva.d?id=62187757"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai"@
"B. Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai"@
"Barackas Obama"@lt	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@

5.3 pav. SPARQL užklauso rezultatai ieškant individo „Obama“

Pagal DISTINCT raktažodį (žr. žemiau pateiktą SPARQL užklausą) atrenkami tik unikalūs įrašai. Jeigu Barackas Obama buvo tame pačiame straipsnyje ir ta pačia žodžio forma paminėtas kelis kartus (nuorodos į dokumentą sutaps), tai ši užklausa rodytų tik vieną įrašą, kuriame buvo paminėtas Barackas Obama su nuoroda į tą patį vieną dokumento URL.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```

PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT DISTINCT ?found_label ?uri
WHERE {
  ?subject a ag:person .
  ?subject an:label_lemma ?label_lemma .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i")) .
  ?subject ag:is_similar_to ?merged .
  ?found ag:is_similar_to ?merged .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
}

```

found_label	uri
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva.d?id=62187757"@
"Barackas Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai"@
"B. Obama"@lt	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@
"Barackas Obama"@lt	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801"@
"B. Obama"@lt	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801"@

5.4 pav. SPARQL užklauso rezultatai ieškant individo „obama“ naudojant DISTINCT raktąžodį

Eksperimento metu buvo peržiūrėti 5 straipsniai, kuriuose paminėtas ontologijos individas Obama.

1. <http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva.d?id=62187757>
2. <http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu>
3. http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_turetu_vykti_paraleliai
4. <http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu>
5. <http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801>

Žemiau pateikta užklausa pirmame straipsnyje grąžina asmenų, rastų konkrečiame straipsnyje, vardus su to straipsnio ID ir URL (žr. 5.5 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
PREFIX ann: <http://semantika.lt/ns/Annotations#>
SELECT ?found_label ?id ?uri
WHERE {
  ?document a sem:document .
  ?document sem:document_ID ?id .
  FILTER( regex(?id, "88dd")) .
  ?document sem:document_URI ?uri .
  FILTER( regex(?uri, "http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva")) .
  ?document sem:refers_to__object ?found .
  ?found a ag:person .
  ?found sem:trusted ?trusted .
  FILTER( regex(?trusted, "^untrusted")) .
  ?found rdfs:label ?found_label .
}

```

found_label	id	uri
"B. Obama"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania
"Algirdas Butkevičius"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania
"Chuckas Hagelis"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania
"A. Butkevičius"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania
"B. al Assada"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania
"Barackas Obama"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania
"Basharo al Assado"@lt	"52bc7419ab9033d40a0088dd"@	"http://www.delfi.lt/news/daily/lithuania

5.5 pav. Pirmame straipsnyje paminėti ontologijos individai

Ontologijos individas Barackas Obama ir individas B. Obama nagrinėjamos ontologijos konkrečiame straipsnyje buvo rasti po 1 kartą (žr. 5.6 pav.). Jie buvo apjungti, kaip panašūs ontologijos individai į Obama.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i")) .
  ?subject ag:is_similar_to ?merged .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "52bc7419ab9033d40a0088dd")) }

```

found_label	merged_label	uri	id
"B. Obama"@lt	"b. obama"@	"http://www.delfi.lt/news/daily/lithuania/a-butkevic"	"52bc7419ab9033d40a0088dd"@
"Barackas Obama"@lt	"Barackas obama"@	"http://www.delfi.lt/news/daily/lithuania/a-butkevic"	"52bc7419ab9033d40a0088dd"@

5.6 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama

Ontologijos individas Algirdas Butkevičius nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą, o individas A. Butkevičius - 5 kartus (žr. 5.7 pav.). Šie individai buvo apjungti, kaip panašūs į ontologijos individą Algirdas.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "algirdas", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .

```

```
?document sem:document_URI ?uri .
?document sem:document_ID ?id . FILTER( regex(?id, "52bc7419ab9033d40a0088dd")) }
```

found_label	merged_label	uri	id
"Algirdas Butkevičius"@lt	"algirdas butkevičius"@	"http://www.delfi.lt/news/daily/lithuania/a-bc"52bc7419ab9033d40a0088dd"@	
"A. Butkevičius"@lt	"algirdas butkevičius"@	"http://www.delfi.lt/news/daily/lithuania/a-bc"52bc7419ab9033d40a0088dd"@	
"A. Butkevičius"@lt	"algirdas butkevičius"@	"http://www.delfi.lt/news/daily/lithuania/a-bc"52bc7419ab9033d40a0088dd"@	
"A. Butkevičius"@lt	"algirdas butkevičius"@	"http://www.delfi.lt/news/daily/lithuania/a-bc"52bc7419ab9033d40a0088dd"@	
"A. Butkevičius"@lt	"algirdas butkevičius"@	"http://www.delfi.lt/news/daily/lithuania/a-bc"52bc7419ab9033d40a0088dd"@	
"A. Butkevičius"@lt	"algirdas butkevičius"@	"http://www.delfi.lt/news/daily/lithuania/a-bc"52bc7419ab9033d40a0088dd"@	

5.7 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Algirdas

Ontologijos individas Chuckas Hagelis nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.8 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "hagelis", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "52bc7419ab9033d40a0088dd")) }
```

found_label	merged_label	uri	id
"Chuckas Hagelis"@lt	"Chuckas Hagelis"@	"http://www.delfi.lt/news/daily/lithuania"52bc7419ab9033d40a0088dd"@	

5.8 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Hagelis

Ontologijos individas Basharo al Assado nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą, taip pat 1 kartą rastas ir individas B. al Assadą (žr. 5.9 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "assadas", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "52bc7419ab9033d40a0088dd")) }
```

found_label	merged_label	uri	id
"B. al Assadą"@lt	"b. al assadas"@	"http://www.delfi.lt/news/daily/lithuania/a-butkevici"52bc7419ab9033d40a0088dd"@	
"Basharo al Assado"@lt	"Basharo al assadas"@	"http://www.delfi.lt/news/daily/lithuania/a-butkevici"52bc7419ab9033d40a0088dd"@	

5.9 pav. Pirmame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Assadas

Žemiau pateikta užklausa antrame straipsnyje gražina asmenų, rastų konkrečiame straipsnyje, vardus su to straipsnio ID ir URL (žr. 5.10 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
PREFIX ann: <http://semantika.lt/ns/Annotations#>
SELECT ?found_label
WHERE {
  ?document a sem:document .
  ?document sem:document_ID ?id .
  FILTER( regex(?id, "53553a4c2ef9255e0a046b15")) .
  ?document sem:document_URI ?uri .
  FILTER( regex(?uri, "http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-
biudzetu/")) .
  ?document sem:refers_to__object ?found .
  ?found a ag:person .
  ?found sem:trusted ?trusted .
  FILTER( regex(?trusted, "^untrusted")) .
  ?found rdfs:label ?found_label .
}
```

found_label	id	uri
"B. Obama"@lt	"53553a4c2ef9255e0a046b15"@	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@
"Barackas Obama"@lt	"53553a4c2ef9255e0a046b15"@	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu/"@

5.10 pav. Antrame straipsnyje paminėti ontologijos individai

Ontologijos individas Barackas Obama konkrečiame straipsnyje buvo rastas 1 kartą, 1 kartą buvo rastas ir individas B. Obama (žr. 5.11 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "53553a4c2ef9255e0a046b15")) }
```

found_label	merged_label	uri	id
"B. Obama"@lt	"b. obama"@	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biud"	"53553a4c2ef9255e0a046b15"@
"Barackas Obama"@lt	"Barackas obama"@	"http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biud"	"53553a4c2ef9255e0a046b15"@

5.11 pav. Antrame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama

Žemiau pateikta užklausa trečiame straipsnyje gražina asmenų, rastų konkrečiame straipsnyje, vardus su to straipsnio ID ir URL (žr. 5.12 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
PREFIX ann: <http://semantika.lt/ns/Annotations#>
SELECT ?found_label ?id ?uri
WHERE {
  ?document a sem:document .
  ?document sem:document_ID ?id .
  FILTER( regex(?id, "a8e") ) .
  ?document sem:document_URI ?uri .
  FILTER(
    regex(?uri,
"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_diskusijos_del_prism_ir_prekybos_
_turetu_vykti_pareleliai")) .
  ?document sem:refers_to__object ?found .
  ?found a ag:person .
  ?found sem:trusted ?trusted .
  FILTER( regex(?trusted, "^untrusted") ) .
  ?found rdfs:label ?found_label .
}

```

found_label	id	uri
"Ericas Holderis"@lt	"5313d5a74d4b5b7e09022a8e"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_di:
"D. Grybauskaitė"@lt	"5313d5a74d4b5b7e09022a8e"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_di:
"Dalia Grybauskaitė"@lt	"5313d5a74d4b5b7e09022a8e"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_di:
"Angela Merkel"@lt	"5313d5a74d4b5b7e09022a8e"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_di:
"Barackas Obama"@lt	"5313d5a74d4b5b7e09022a8e"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente_di:

5.12 pav. Trečiame straipsnyje paminėti ontologijos individai

Ontologijos individas Barackas Obama nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.13 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i") ) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "5313d5a74d4b5b7e09022a8e")) }

```

found_label	merged_label	uri	id
"Barackas Obama"@lt	"Barackas obama"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente"	"5313d5a74d4b5b7e09022a8e"@

5.13 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama

Ontologijos individas Ericas Holderis nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.14 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>

```

```

PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "ericas", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "5313d5a74d4b5b7e09022a8e"))}

```

found_label	merged_label	uri	id
"Ericas Holderis"@lt	"Ericas Holderis"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente"5313d5a74d4b5b7e09022a8e"@	

5.14 pav. Trečiamie straipsnyje rasti ir apjungti ontologijos individai, panašūs į Ericas

Ontologijos individas Dalia Grybauskaitė konkrečiame straipsnyje buvo rastas 1 kartą, 1 kartą buvo rastas ir individas D. Grybauskaitė (žr. 5.15 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "dalia", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "5313d5a74d4b5b7e09022a8e"))}

```

found_label	merged_label	uri	id
"D. Grybauskaitė"@lt	"dalia grybauskaitė"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezid"5313d5a74d4b5b7e09022a8e"@	
"Dalia Grybauskaitė"@lt	"dalia grybauskaitė"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezid"5313d5a74d4b5b7e09022a8e"@	

5.15 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Dalia

Ontologijos individas Angela Merkel nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.16 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "merkel", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .

```

```
?document sem:refers_to__object ?found .
?document sem:document_URI ?uri .
?document sem:document_ID ?id . FILTER( regex(?id, "5313d5a74d4b5b7e09022a8e")) }
```

found_label	merged_label	uri	id
"Angela Merkel"@lt	"Angela merkel"@	"http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezli"	"5313d5a74d4b5b7e09022a8e"@

5.16 pav. Trečiame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Merkel

Žemiau pateikta užklausa ketvirtame straipsnyje grąžina asmenų, rastų konkrečiame straipsnyje, vardus su to straipsnio ID ir URL (žr. 5.17 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
PREFIX ann: <http://semantika.lt/ns/Annotations#>
SELECT ?found_label ?id ?uri
WHERE {
  ?document a sem:document .
  ?document sem:document_ID ?id .
  FILTER( regex(?id, "0d3")) .
  ?document sem:document_URI ?uri .
  FILTER( regex(?uri, "http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-
apie-galimus-maskvos-planus-del-baltijos-saliu")) .
  ?document sem:refers_to__object ?found .
  ?found a ag:person .
  ?found sem:trusted ?trusted .
  FILTER( regex(?trusted, "^untrusted")) .
  ?found rdfs:label ?found_label .
}
```

found_label	id	uri
"Linas Linkevičius"@lt	"533c68512ef9255e0a0270d3"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@
"L. Linkevičius"@lt	"533c68512ef9255e0a0270d3"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@
"Barackas Obama"@lt	"533c68512ef9255e0a0270d3"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@
"Juozas Olekas"@lt	"533c68512ef9255e0a0270d3"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu"@

5.17 pav. Ketvirtame straipsnyje paminėti ontologijos individai

Ontologijos individas Barackas Obama nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.18 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "533c68512ef9255e0a0270d3")) }
```

found_label	merged_label	uri	id
"Barackas Obama"@lt	"Barackas obama"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-p"	"533c68512ef9255e0a0270d3"@

5.18 pav. Ketvirtame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama

Ontologijos individas Linas Linkevičius konkrečiame straipsnyje buvo rastas 1 kartą, 1 kartą buvo rastas ir individas L. Linkevičius (žr. 5.19 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "linas", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "533c68512ef9255e0a0270d3"))}
```

found_label	merged_label	uri	id
"Linas Linkevičius"@lt	"linas linkevičius"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongr"533c68512ef9255e0a0270d3"@	
"L. Linkevičius"@lt	"linas linkevičius"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongr"533c68512ef9255e0a0270d3"@	

5.19 pav. Ketvirtame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Linas

Ontologijos individas Juozas Olekas nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.20 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "olekas", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "533c68512ef9255e0a0270d3"))}
```

found_label	merged_label	uri	id
"Juozas Olekas"@lt	"juozas olekas"@	"http://www.alfa.lt/straipsnis/15847692/jav-kongresas-per"533c68512ef9255e0a0270d3"@	

5.20 pav. Ketvirtame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Olekas

Žemiau pateikta užklausa penktame straipsnyje grąžina asmenų, rastų konkrečiame straipsnyje, vardus su to straipsnio ID ir URL (žr. 5.21 pav.).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
```

```

PREFIX sem: <http://semantika.lt/ns/SemLT#>
PREFIX ann: <http://semantika.lt/ns/Annotations#>
SELECT ?found_label ?id ?uri
WHERE {
  ?document a sem:document .
  ?document sem:document_ID ?id .
  FILTER( regex(?id, "fb9")) .
  ?document sem:document_URI ?uri .
  FILTER( regex(?uri, "http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-
mano-kad-bobama-gali-buti-antikristas")) .
  ?document sem:refers_to__object ?found .
  ?found a ag:person .
  ?found sem:trusted ?trusted .
  FILTER( regex(?trusted, "^untrusted")) .
  ?found rdfs:label ?found_label .
}

```

found_label	id	uri
"Barackas Obama"@lt	"530f331c4d4b5b7e0901afb9"@	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikie
"B. Obama"@lt	"530f331c4d4b5b7e0901afb9"@	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikie
"Paulas McCartney"@lt	"530f331c4d4b5b7e0901afb9"@	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikie

5.21 pav. Penktame straipsnyje paminėti ontologijos individai

Ontologijos individas Barackas Obama ir individas B. Obama nagrinėjamos ontologijos konkrečiame straipsnyje buvo rasti po 1 kartą (žr. 5.22 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .
  ?subject rdfs:label ?label .
  FILTER( regex(?label, "obama", "i")) .
  ?found ag:is_similar_to ?merged .
  ?merged an:label_lemma ?merged_label .
  ?found rdfs:label ?found_label .
  ?document sem:refers_to__object ?found .
  ?document sem:document_URI ?uri .
  ?document sem:document_ID ?id . FILTER( regex(?id, "530f331c4d4b5b7e0901afb9"))}

```

found_label	merged_label	uri	id
"Barackas Obama"@lt	"Barackas obama"@	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikie	"530f331c4d4b5b7e0901afb9"@
"B. Obama"@lt	"b. obama"@	"http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikie	"530f331c4d4b5b7e0901afb9"@

5.22 pav. Penktame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Obama

Ontologijos individas Paulas McCartney nagrinėjamos ontologijos konkrečiame straipsnyje buvo rastas 1 kartą (žr. 5.23 pav.).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX mo: <http://semantika.lt/ns/MergedObjects#>
PREFIX ag: <http://semantika.lt/ns/Agents#>
PREFIX an: <http://semantika.lt/ns/Annotations#>
PREFIX sem: <http://semantika.lt/ns/SemLT#>
SELECT ?found_label ?merged_label ?uri ?id
WHERE {
  ?subject a ag:person .

```

```
?subject rdfs:label ?label .
FILTER( regex(?label, "paulas", "i")) .
?found ag:is_similar_to ?merged .
?merged an:label_lemma ?merged_label .
?found rdfs:label ?found_label .
?document sem:refers_to__object ?found .
?document sem:document_URI ?uri .
?document sem:document_ID ?id . FILTER( regex(?id, "530f331c4d4b5b7e0901afb9")) }
```

found_label	merged_label	uri	id
"Paulas McCartney"@lt	"paulas McCartney"@	"http://www.delfi.lt/news/daily/world/kas-ketvirta:"530f331c4d4b5b7e0901afb9"@	

5.23 pav. Penktame straipsnyje rasti ir apjungti ontologijos individai, panašūs į Paulas

Žmogus, peržiūrėjęs nurodytus straipsnius, gavo štai tokius rezultatus:

Pirmame straipsnyje (<http://www.delfi.lt/news/daily/lithuania/a-butkevicius-apie-grasinimus-kad-rusija-uzpuls-lietuva.d?id=62187757>) žmogus atpažino šiuos individus:

- B. Obama
- Barackas Obama
- Algirdas Butkevičius
- A. Butkevičius
- A. Butkevičius
- A. Butkevičius
- A. Butkevičius
- A. Butkevičius
- Chuckas Hagelis
- B. al Assada
- B. al Assado
- Basharo al Assado

Antrame straipsnyje (<http://iq.lt/ekonomika/jav-politikai-kausis-del-dvieju-biudzetu>) žmogus atpažino šiuos individus:

- B. Obama
- Barackas Obama

Trečiame straipsnyje (<http://www.lrt.lt/naujienos/lietuvoje/2/20995/prezidente-diskusijos-del-prism-ir-prekybos-turetu-vykti-paraleliai>) žmogus atpažino šiuos individus:

- Ericas Holderis
- D. Grybauskaitė
- D. Grybauskaitė
- D. Grybauskaitė
- D. Grybauskaitė
- D. Grybauskaitė
- Dalia Grybauskaitė
- Angela Merkel
- Barackas Obama

Ketvirtame straipsnyje (<http://www.alfa.lt/straipsnis/15847692/jav-kongresas-perspeja-apie-galimus-maskvos-planus-del-baltijos-saliu>) žmogus atpažino šiuos individus:

- Linas Linkevičius
- L. Linkevičius
- Barackas Obama
- Juozas Olekas

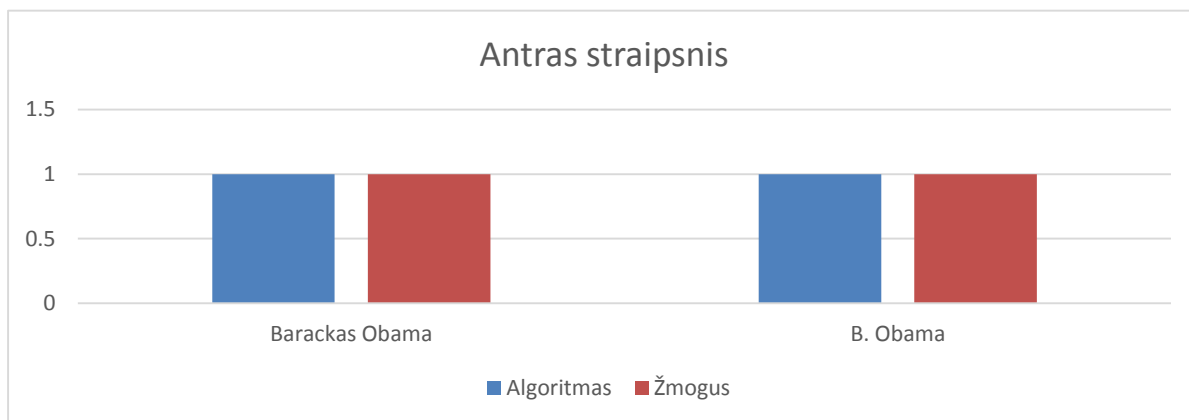
Na, o penktame straipsnyje (<http://www.delfi.lt/news/daily/world/kas-ketvirtas-amerikietis-mano-kad-bobama-gali-buti-antikristas.d?id=61055801>) žmogus atpažino šiuos individus:

- Barackas Obama
- B.Obama
- B. Obama
- B. Obama
- B. Obama
- Paulas McCartney

Žemiau pateikiama lentelė, kurioje matomi žmogaus ir teksto anotatorių rasti asmenų paminėjimai 5 straipsniuose.

5.1 lentelė. Žmogaus ir teksto anotatorių rastų asmenų paminėjimų straipsniuose lyginamoji lentelė

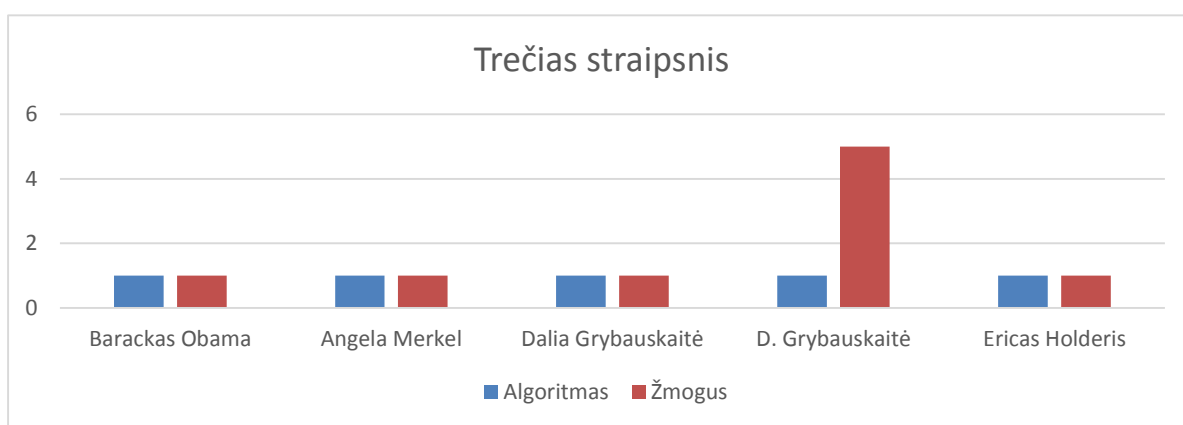
Eil. Nr.	Žmogus	Teksto anotatoriai
1 straipsnis		
1.	B. Obama	B. Obama
2.	Algirdas Butkevičius	Algirdas Butkevičius
3.	Chuckas Hagelis	Chuckas Hagelis
4.	A. Butkevičius	A. Butkevičius
5.	A. Butkevičius	A. Butkevičius
6.	A. Butkevičius	A. Butkevičius
7.	A. Butkevičius	A. Butkevičius
8.	A. Butkevičius	A. Butkevičius
9.	B. al Assadą	B. al Assadą
10.	Barackas Obama	Barackas Obama
11.	Basharo al Assado	Basharo al Assado
12.	B. al Assado	
<p style="text-align: center;">5.24 pav. Pirmo straipsnio lyginamoji lentelė</p>		
2 straipsnis		
1.	B. Obama	B. Obama
2.	Barackas Obama	Barackas Obama



5.25 pav. Antro straipsnio lyginamoji lentelė

3 straipsnis

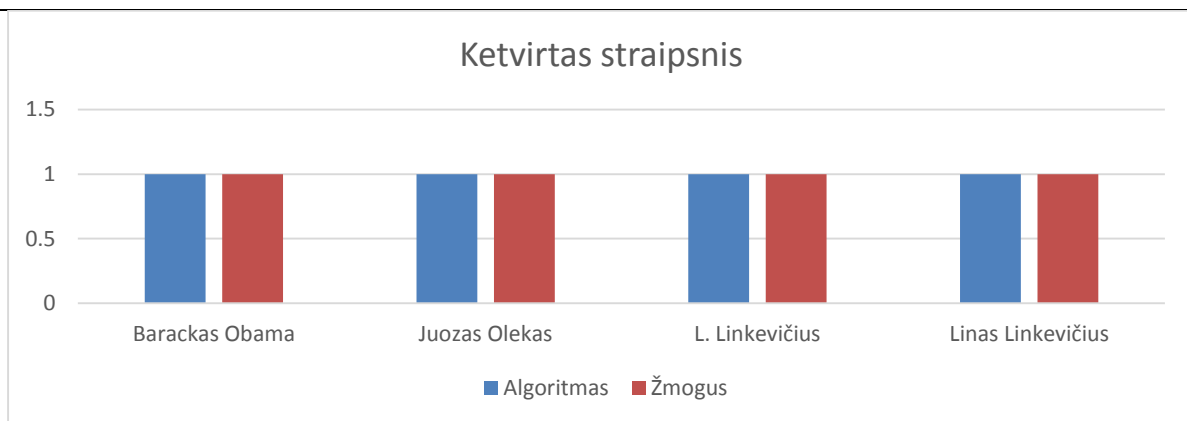
1.	Ericas Holderis	Ericas Holderis
2.	D. Grybauskaitė	D. Grybauskaitė
3.	Dalia Grybauskaitė	Dalia Grybauskaitė
4.	Angela Merkel	Angela Merkel
5.	Barackas Obama	Barackas Obama
6.	D. Grybauskaitė	
7.	D. Grybauskaitė	
8.	D. Grybauskaitė	
9.	D. Grybauskaitė	



5.26 pav. Trečio straipsnio lyginamoji lentelė

4 straipsnis

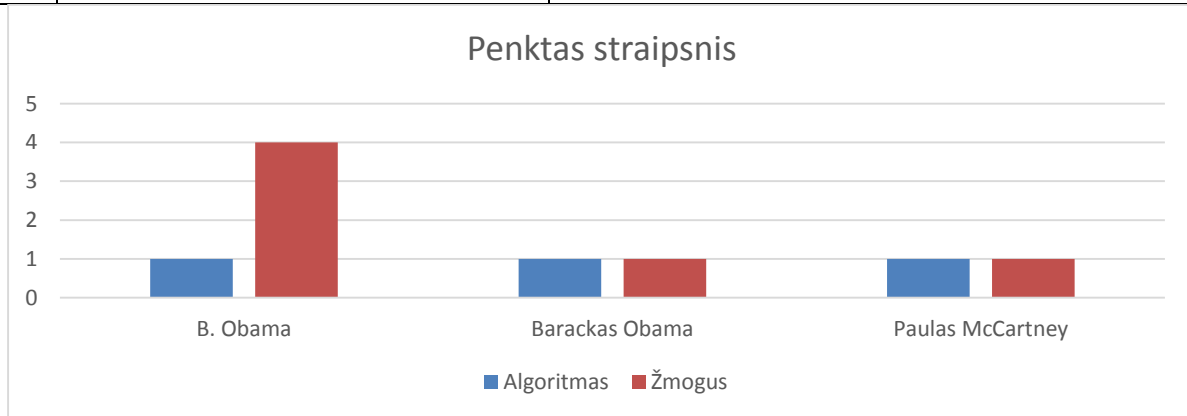
1.	Linas Linkevičius	Linas Linkevičius
2.	L. Linkevičius	L. Linkevičius
3.	Barackas Obama	Barackas Obama
4.	Juozas Olekas	Juozas Olekas



5.27 pav. Ketvirto straipsnio lyginamoji lentelė

5 straipsnis

1.	Barackas Obama	Barackas Obama
2.	B.Obama	B. Obama
3.	Paulas McCartney	Paulas McCartney
4.	B. Obama	
5.	B. Obama	
6.	B. Obama	



5.28 pav. Penkto straipsnio lyginamoji lentelė

1, 3, 5 straipsniuose teksto analizatoriai blogai atpažino paminėtus asmenis. Dėl šios priežasties tekste paminėtų asmenų skaičius nesutampa su ontologijoje išsaugotų individų skaičiumi. Nekokybiški teksto analizės rezultatai dažnai yra semantinio anotavimo komponento taisyklių ir leksinės analizės komponentų ribotumas, jų netobulumas.

Siekiant vertinti algoritmo kokybę buvo palyginti žmogaus ir algoritmo rasti ir apjungti panašūs individai (žr. 5.2 lentelė).

5.2 lentelė. Žmogaus ir algoritmo rastų ir apjungtų panašių ontologijos individų lyginamoji lentelė

Eil. Nr.	Žmogus	Algoritmas
1 straipsnis		
1.	Barackas Obama <ul style="list-style-type: none"> Barackas Obama B. Obama 	b. obama <ul style="list-style-type: none"> B. Obama Barackas obama <ul style="list-style-type: none"> Barackas Obama
2.	Algirdas Butkevičius <ul style="list-style-type: none"> Algirdas Butkevičius 	algirdas butkevičius <ul style="list-style-type: none"> Algirdas Butkevičius

Eil. Nr.	Žmogus	Algoritmas
	<ul style="list-style-type: none"> A. Butkevičius 	<ul style="list-style-type: none"> A. Butkevičius
3.	Chuckas Hagelis <ul style="list-style-type: none"> Chuckas Hagelis 	Chuckas Hagelis <ul style="list-style-type: none"> Chuckas Hagelis
4.	Basharo al Assadas <ul style="list-style-type: none"> Basharo al Assado B. al Assada B. al Assado 	b. al assadas <ul style="list-style-type: none"> B. al Assada Basharo al assadas <ul style="list-style-type: none"> Basharo al Assado
2 straipsnis		
1.	Barackas Obama <ul style="list-style-type: none"> Barackas Obama B. Obama 	b. obama <ul style="list-style-type: none"> B. Obama Barackas obama <ul style="list-style-type: none"> Barackas Obama
3 straipsnis		
1.	Ericas Holderis <ul style="list-style-type: none"> Ericas Holderis 	Ericas Holderis <ul style="list-style-type: none"> Ericas Holderis
2.	Dalia Grybauskaitė <ul style="list-style-type: none"> Dalia Grybauskaitė D. Grybauskaitė 	dalia grybauskaitė <ul style="list-style-type: none"> D. Grybauskaitė Dalia Grybauskaitė
3.	Angela Merkel <ul style="list-style-type: none"> Angela Merkel 	Angela merkel <ul style="list-style-type: none"> Angela Merkel
4.	Barackas Obama <ul style="list-style-type: none"> Barackas Obama 	Barackas obama <ul style="list-style-type: none"> Barackas Obama
4 straipsnis		
1.	Linas Linkevičius <ul style="list-style-type: none"> Linas Linkevičius L. Linkevičius 	linas linkevičius <ul style="list-style-type: none"> Linas Linkevičius L. Linkevičius
2.	Barackas Obama <ul style="list-style-type: none"> Barackas Obama 	Barackas obama <ul style="list-style-type: none"> Barackas Obama
3.	Juozas Olekas <ul style="list-style-type: none"> Juozas Olekas 	juozas olekas <ul style="list-style-type: none"> Juozas Olekas
5 straipsnis		
1.	Barackas Obama <ul style="list-style-type: none"> Barackas Obama B.Obama 	b. obama <ul style="list-style-type: none"> B. Obama Barackas obama <ul style="list-style-type: none"> Barackas Obama
2.	Paulas McCartney <ul style="list-style-type: none"> Paulas McCartney 	paulas McCartney <ul style="list-style-type: none"> Paulas McCartney

Ontologijos individų B. Obama ir Barackas Obama merged_label skirtinga, nes algoritmas šiuos individus sujungė su skirtingais merged_person. Ta pati situacija ir su ontologijos individualais B. al Assada ir Basharo al Assado.

5.3 lentelėje pateikti algoritmo teisingai, klaidingai nustatyti ir nesurasti panašumai.

5.3 lentelė. Algoritmo teisingai, klaidingai nustatytų ir nesurastų panašumų lentelė

Straipsnio Nr.	Teisingai nustatyti panašumai	Klaidingai nustatyti panašumai	Nesurasti panašumai / Neatpažinta
1	4	0	3
2	1	0	1

Straipsnio Nr.	Teisingai nustatyti panašumai	Klaidingai nustatyti panašumai	Nesurasti panašumai / Neatpažinta
3	4	0	0
4	3	0	0
5	2	0	1
Σ	14	0	5

- Teisingai nustatytų panašumų (truePositives) skaičius: 14
- Klaidingai nustatytų panašumų (falsePositives) (klaidingų elementų, palaikytų teisingais (panašiais)) skaičius: 0
- Nesurasti panašumai (falseNegatives) (praleistų elementų (klaidingai nepalaikytų panašiais)) skaičius ontologijoje: 5
- Tikslumas: teisingai nustatytų panašumų santykis su visais (teisingais ir neteisingais) nustatytais panašumais :
 $(\text{truePositives} / (\text{truePositives} + \text{falsePositives}))$.
 $(14 / (14 + 0))=1$
- Atkūrimas (teisingai nustatytų panašumų santykis su visais (nustatytais ir nenustatytais) panašumais):
 $(\text{truePositives} / (\text{truePositives} + \text{falseNegatives}))$. [17]
 $(14 / (14 + 5))=0,74$
- Kokybė: harmonizuota tikslumo ir atkūrimo išraiška
 $(2 * (\text{tikslumas} * \text{atgaminimas}) / (\text{tikslumas} + \text{atgaminimas}))$. [17]
 $(2 * (1 * 0,74) / (1 + 0,74))=0,85$

Eksperimento rezultatai atitiko kokybės kriterijus 85%.

5.3. Sprendimo taikymo rekomendacijos

Sukurtas sprendimas lengvai pritaikomas tokio tipo ontologijai, kurios struktūroje yra label_lemma (bendrinė žodžio forma), naudojant label_lemma's bus sumažinti ontologijos individų pasikartojimai, sujungiant juos pagal panašumą ir taip siekiant pateikti konkrečią, struktūrizuotą informaciją.

6. REZULTATŲ APIBENDRINIMAS IR IŠVADOS

1. Šio Magistrinio baigiamojo darbo tikslas – sudaryti galimybę automatiškai nustatyti panašius ontologijos individus, juos susieti ir tokiu būdu sumažinti informacijos ontologijoje dubliavimą – pasiektas.
2. Išanalizuoti metodai – Levenshtein'o atstumas, Damerau–Levenshtein'o atstumas ir Hamming'o atstumas, kurie skirti individų panašumui surasti. Analizė parodė, kad esami sprendimai turi trūkumų.
3. Sukurtos dvi ontologijų suliejimo taisyklės: suliejimas pagal vardą ir pavardę, kai pavardė pilna, o vardas sutrumpintas ir suliejimas, kai sutampa individų vardai ir pavardės.
4. Sukurta nauja schema *MergedObjects*, kurioje yra esybė *merged_person* ir papildyta *Agents* schema, į kurią įdėta savybė *is_similar_to*, tam kad būtų susiejami panašūs ontologijos individai.
5. Parašytos SPARQL užklausos, kurias įvykdžius išvedami panašūs asmenys su pilnais vardais ir nuorodomis į dokumentą.
6. Sukurtas algoritmas sudaro galimybę automatiškai nustatyti panašius ontologijos individus ir juos susieti.
7. Sukurtas sprendimas lengvai pritaikomas tokio tipo ontologijai, kurios struktūroje yra *label_lemma* (bendrinė žodžio forma).

7. LITERATŪRA

- [1] R. J. Rimutė Aranauskienė, „LIETUVIŲ KALBOS KURSAS 5–6 KLASĖMS,“ [Tinkle]. Available: http://lietuviu5-6.mkp.emokykla.lt/lt/mo/zinynas/kalbos_pazinimas_lietuviu_kalbos_zodziu_daryba_kaityba_sandara_morfologija/. [Kreiptasi 04 2015].
- [2] NOY N. F., McGUINNESS D. L., *Ontology Development : A Guide to Creating Your First Ontology*..
- [3] N. Guarino, „Formal Ontology and Information Systems,“ Italy, 1998, pp. 3-15.
- [4] M. Uschold, M. King, S. Moralee, Y. Zorgios, „The Enterprise Ontology,“ įtraukta *The Knowledge Engineering Review*, AIAI-TR-195.
- [5] BPMI, „BPMI,“ 02 01 2008. [Tinkle]. Available: www.bpmi.org/. [Kreiptasi 06 2014].
- [6] A. Čaplinskas, A. Lupeikienė, Unified Enterprise engineering environment: ontological point of view, Kluwer: BalticDB&IS*2002 proceedings, 2002.
- [7] Benjamins, V. R., „Overview of knowledge sharing and reuse components: ontologies and problem – solving methods,“ įtraukta *IJCAI–99 workshop on Ontologies and Problem–Solving Methods: Lessons Learned and Future Trends*, Stockholm, Sweden, 1999.
- [8] Wikipedia, „Ontology (information science),“ [Tinkle]. Available: http://en.wikipedia.org/wiki/Ontology_%28information_science%29. [Kreiptasi 11 2014].
- [9] Wikipedia, „Hamming_distance,“ [Tinkle]. Available: http://en.wikipedia.org/wiki/Hamming_distance. [Kreiptasi 04 2015].
- [10] Wikipedia, „Wikipedia,“ Levenshtein distance, [Tinkle]. Available: http://en.wikipedia.org/wiki/Levenshtein_distance. [Kreiptasi 03 2015].
- [11] Oommen, B. J.; Loke, R. K. S., „Pattern recognition of strings with substitutions, insertions, deletions and generalized transpositions,“ *Pattern Recognition*, t. 30, nr. 5, p. 789–800, 1997.
- [12] V., Gregory, „Spelling-error tolerant, order-independent pass-phrases via the Damerau–Levenshtein string-edit distance metric,“ įtraukta *the Fifth Australasian Symposium on ACSW Frontiers : 2007, Ballarat, Australia, January 30 - February 2, 2007*, Darlinghurst, Australia, 2007.
- [13] Fred J. Damerau, „A technique for computer detection and correction of spelling errors,“ *Communications of the ACM*, t. 7, nr. 3, p. 171–176, 1964.
- [14] Majorek, Karolina A.; Dunin-Horkawicz, Stanisław, „The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification,“ *Nucleic Acids Research*, t. 42, nr. 7, p. 4160–4179, 2013.
- [15] Levenshtein, Vladimir I., „Binary codes capable of correcting deletions, insertions, and reversals,“ *Soviet Physics Doklady*, t. 10, nr. 8, p. 707–710, 1966.
- [16] V. I. LEVENSHTAIN, Binary codes capable of correcting deletions, insertions, and reversals, *Doklady Akademii Nauk*, 1965.
- [17] Wikipedia, „Sensitivity_and_specificity,“ [Tinkle]. Available: http://en.wikipedia.org/wiki/Sensitivity_and_specificity. [Kreiptasi 12 2014].